Graduate Theses, Dissertations, and Problem Reports

2018

Reducing Spatial Stochastic Models of Membrane Receptors to Approximately Equivalent Chemical Reaction Networks through Coarse Graining

Christopher Short

Follow this and additional works at: https://researchrepository.wvu.edu/etd

Reducing Spatial Stochastic Models of Membrane Receptors to
Approximately Equivalent Chemical Reaction Networks through
Coarse Graining

Christopher Short

Dissertation submitted to the
Eberly College of Arts and Sciences
at West Virginia University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in
Mathematics

Ádam M. Halász, Ph.D, Chair
Harvey Diamond, Ph.D
Casian Pantea, Ph.D
Adrian Tudoraşcu, Ph.D
Jeremy S. Edwards, Ph.D

Department of Mathematics

Morgantown, WV
2018

Keywords: Cell Signal Initiation, Coarse Graining, VEGF, EGF,
Spatial Models

ABSTRACT

Reducing Spatial Stochastic Models of Membrane Receptors to Approximately Equivalent Chemical Reaction Networks through Coarse Graining

Christopher Todd Short

Computational models are necessary for understanding how biological functionality emerges from a complex network of molecular interactions such as in cell signaling. Due to the complexity of biological systems, the normal scientific method of hypothesizing then testing predictions is becoming increasingly difficult. The large number of processes limits the level of modeling detail; signaling models typically adopt a non-spatial representation where each molecular process is characterized by a few parameters.

However, modern microscopic imaging of receptors on cell membranes reveals an intricate structure of microdomains. Receptors, such as epidermal growth factor receptor (EGFR), vascular endothelial growth factor receptor (VEGFR), and others, tend to be grouped in the microdomains as clusters that range from a few to hundreds of biomolecules. While the origin of these clusters is not well understood, a likely explanation is the existence of microdomains with an affinity for the receptors. Using this hypothesis, we can ignore the underlying cause of the microdomains. The size and geometry of these domains can be inferred directly from microscopy; however, the relevant physical properties can only be verified through simulations. In this thesis, I propose a flexible approach to performing such simulations in a coarse grained model that is validated through solving the differential equations when possible and through equilibrium calculations when not.

Due to the non-trivial nature of these cell membrane features, fully spatial models need to be used to address these issues. However, fully spatial models are computationally intense and little insight can be gained from them, because of this I propose a method to construct the well-mixed model from the spatial one. The primary issue is the difficulty of extracting the correct kinetic coefficients and that limits the predictive power of spatial models due to the inherent challenges of estimating dynamical parameters. I will use the

Metropolis-Hastings Algorithm to extract these parameters from the fully spatial simulations. I use the spatial model as an intermediary step because the spatial simulations can be matched to experimental techniques that provide molecular level resolution, such as single particle tracking (SPT) data. I will then discuss issues that emerge from a baseline comparison between spatial and non-spatial simulations of a simple reversible dimerization process; the spatial simulation employed an algorithm similar to Smoldyn.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

Mathematical, Quantitative, or Systems Biology all refer to an effort to introduce a mathematical point of view in the practice of Biology, in hopes of some day matching the success of the 400-year old alliance of Mathematics with Physics.

The vital functions of a cell emerge from its component molecules and their transformations. To develop a predictive understanding of exactly how this occurs, we have to build an inventory of the elementary (molecular) components, the processes they participate in and the physical laws that govern those processes; then put them all together in a mathematical[1] model that captures the complete state of the cell and emulates all the processes and their interdependence.

The potential impact of this level of understanding is hard to fathom; continuing the analogy with Physics, consider that much of the technology that came about after 1700, from the steam engine to computers and space travel, was applications of Physics and Chemistry in various engineering settings. Things could be designed and optimized because the mathematical framework provided a way to summarize and analyze experimental facts, and to predict the outcomes of experiments never done before.

For cell biology, the most obvious application is health and disease; a complete understanding of the biochemistry of our cells would let us under-

---

[1]In the general sense of a system of mathematical objects that represent the states and transformations of the real system.

stand exactly how an adverse health condition emerges, and would also help us figure out how to stop it from becoming a problem. The most important obstacles to this vision are that (1) it is difficult to identify each relevant molecular species or process, and even harder to quantify their physical laws; (2) even in a single cell, there are a very large number of different molecular species and processes at work, some of them involving no more than a handful of molecules. However, developments in the past few decades have been promising. In addition to DNA sequencing and related methods, microscopic imaging and labeling techniques can provide in vivo information on individual molecules. Advances in computer technology made it possible to automatically handle and reason about vast amounts of data. As biochemistry and its techniques advanced, the amount of data and ergo our understanding of cellular biology will also advance which will allow for an ever increasing level of detail in our cellular models. Many cellular models are very complicated and gaining insight from these models requires computer simulations. For these models to be useful they require correct model parameters that match experimental data. This thesis addresses some of these issues.

In this thesis I will consider two chemical reaction networks for cell signaling, specifically starting with vascular endothelial growth factor (VEGF) then moving to epidermal growth factor (EGF), although the methods described can be applied to any chemical reaction network where the reactions occur in subdomains with movement between the subdomains. This thesis focuses on VEGF and EGF because of medical relevance which will be described later. We begin with a brief discussion of different mathematical models involved.

We are modeling chemical reaction networks (CRNs). These reactions can be modeled several ways; as ordinary differential equations (ODEs), stochastically as with the Gillespie SSA, or fully spatially. When considering ODEs or SSA, the reactions are treated as mass action. Mass Action Rate Laws are used to construct differential equations from the CRN. Mass action is also the assumption used for construction of the probabilities involved in the Gillespie SSA. Fully spatial models utilize diffusing particles in a space. This utilization allows for a direct comparison to experimental data.

Much cellular communication is accomplished without the signaling molecules directly entering the cell and rather binding to a receptor first for propagation of signal.[27]. The receptor exists on the cellular membrane with intracellular, extracellular and transmembrane domains. The extracellular domain of the receptor include the ligand binding domain as well as domains that en-

able receptor dimerization through the dimerizing arm of the receptor [20]. After being created by the cell and reaching the membrane, the receptor performs a random walk on the cellular membrane attempting to locate its target ligand[27]. When the exterior piece of the receptor becomes bound to its target ligand, the intracellular kinase domain can become activated and specific amino acid residues, such as a tyrosine residue, can be phosphorylated. Once phosphorylated, these residues serve as docking sites for other proteins, which help to further propagate the signal internally.[27]. This process of receptors binding to ligands and transmitting the information into the cell is known as cell signaling, and this can lead to various cellular processes such as cell proliferation, differentiation and survival.

One goal of simulating chemical reaction networks (CRNs) is to have a scalable model of biochemistry employed by cells[37]. Another goal of the simulation process of cell signaling is to predict how a cell will react to a set of external conditions[27]. There are great limitations on simulations because of the sheer number of combinations of substances that can be manufactured by protein interactions, but some of the interactions can be characterized by a set of rules, allowing the computer to generate the species set[61]. However, some simulations can imply astronomical numbers of species[37].

The rules-based approach to the construction of CRN's has been successfully applied to many systems. This is a computer programming approach for constructing the stoichiometric matrices and rate vectors from a set of rules. Receptor-ligand systems of interest include EGF using a graph of the interactions[37]. There has also been success with the trivalent ligand, bivalent receptor model (TLBR), where large aggregates form in the equilibrium model[61]. CRN's generated by rules also have computational limits that are on the order of $log_2 M$ (M is the number of reactions possible) for kinetic Monte Carlo. Much work has been done to lower the limit to the order of the number of rules[61]. There also appears to be more structure to the cellular membrane, which are domains and the heart of this thesis.

1.2 Biology and Experimental Background

1.2.1 Role and Importance of Membrane Receptors

Like all receptors, EGFR and VEGFR are created internally, but migrate to the exterior of the cell. They become bound to the membrane with internal

and external pieces. Then like all receptors, they diffuse across the cellular membrane in an attempt to locate their target ligand. Once the receptor has captured the ligand, in the specific cases of VEGFR and EGFR they must dimerize in order to propagate the signal into the cell. This is not true of all receptor-ligand systems. Once the signal has been transmitted into the cell, the cell will then respond to the signal, leaving a delay between when the receptors first send the signal and the cell responding.

EGFR is very similar to VEGFR in many ways. Both are receptor tyrosine kinases (RTK) receptors and possess a Src Homology 2 (SH2) domain. Both play roles in cancerous development. VEGF promotes angiogenesis. As such, both are subjects for pharmaceutical research. This makes both receptor-ligand systems important for medical science. The starkest contrast between the two systems comes from VEGF ligand being bivalent [21] while EGF ligand is monovalent [39]. Because of this difference, VEGF receptors can experience high dose inhibition while EGF cannot. This means that the higher the concentration of ligand is for VEGF the higher the signal but only to a point. If one fills too many receptors then they cannot find a partner. A diagram of ligand concentration vs. signal would show a max value for the ligand as compared to the signal. This is not the case for EGF. Any diagram showing ligand concentration vs. signal would show a monotonically non-decreasing graph. Moreover that graph would be monotonically increasing until all receptors are filled. High dose inhibition would occur in any receptor-ligand system where the receptor is monovalent with a bivalent ligand, and a lack thereof would also occur when both are monovalent. Also in this paper we will limit the number of receptor and ligand variants we look at due to the combinatorial explosion of numbers of species due to heterodimers. Many receptor-ligand systems have multiple variants and as such share this combinatorial explosion of the number of species.

1.2.2   Microdomains

Experimental observations, such as Single Particle Tracking (SPT) and Transmission Electron Microscopy (TEM), show that the receptors tend to reside in smaller domains of the cell membrane which restrict or hinder the movement of membrane proteins, such as receptors[36]. These domains place the receptors in closer proximity to one and other than if the receptors were distributed throughout the cellular membrane, which should quicken the rate of dimerization for VEGFR. This adds yet another layer of complexity to the
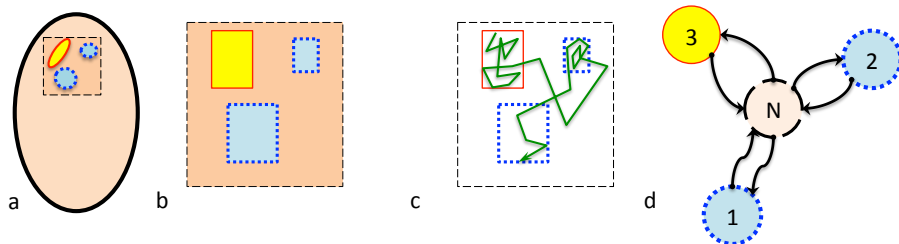
Figure 1.1: Schematic of microdomains located on the cellular membrane and their role in our hypothesis. (a) Microdomains that have an affinity for receptors occupy a small fraction of the cell membrane. Schematically, they may be represented as rectangles (b) whose measurements are extracted from experimental images. (c) Although the receptors can move through the non-attractive regions, they tend to remain in the smaller clusters. (d) In the mesoscopic approach we discuss here, each microdomain, as well as the rest of the membrane (the "normal" region), are represented as well-mixed compartments that may exchange particles.

system as now we need to consider the VEGF network as occurring locally over several microdomains with movement between the domains (See Figure 2). With this scheme, a new species and reaction are created for each species and reaction in the original system for each domain, then movement reactions are added to transport each species across domains. Considering the homodimer case with ten linear locations and movement only between the closest neighbors, we arrive at nearly 200 reactions. This is minor compared to the over 200 reactions for the homodimer EGF case. As such detailed explanations will be presented using VEGF, although the model works for EGF as well.

Due to the sheer size of the VEGF system (which contains 7 species and 7 reversible reactions) including microdomains (the number of species and reactions increases dramatically even with only 2 domain the system has 14 species and 56 reactions), in this thesis, we propose a method to compute the stoichiometric matrices from knowledge of the local system and the movements possible. This approach of combining the molecular components with location information (which acts as a rule as in the rule based approach to create CRNs from a set of rules) has several benefits but none more useful than saving on computational power over a fully spatial model which would be necessary to show this level of detail otherwise. In order to generate the stoichiometry matrices, we treat each location as having a local stoichiometry
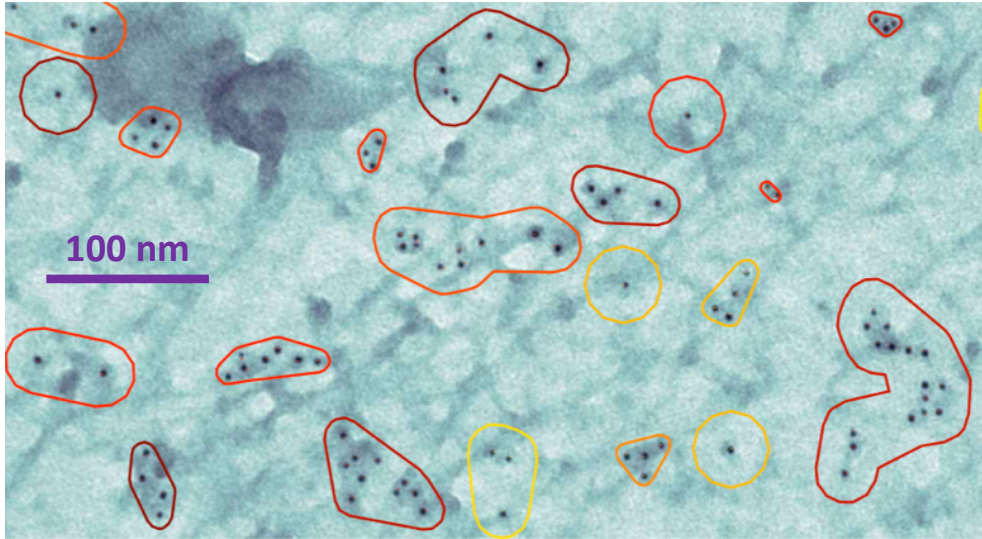
Figure 1.2: Identification of microdomains from microscopy studies. The transmission electron microscopy image on the right was obtained by labeling membrane proteins similar to VEGF receptors with nanometer size gold particles [45].

matrix for both the inputs of the reaction and the outputs. Using the local CRN matrices, larger matrices yielding all molecular reactions are created using the Kronecker Product (will be explained later in the Appendix). Adding transfer information to the total molecular reaction matrices will complete the process and yield the total stoichiometric matrices.

1.2.3 Single Particle Tracking and Transmission Electron Microscopy

I begin with a more detailed discussion of Transmission Electron Microscopy (TEM) imaging. TEM uses nanoscaled probe between $5 - 15nm$ [63]. This method provided the first evidence of receptor subdomains [63]. Once a cell is plated it is then imaged at $30 - 40,000\times$ magnification [63]. Then positions are derived and statistics are used to prove the existence of receptor subdomains [63]. In order to arrive at the proper kinetic coefficients for the parameters of the Well-Mixed model of EGF, we must first consider experimental data in the form of Single Particle Tracking (SPT) data. SPT
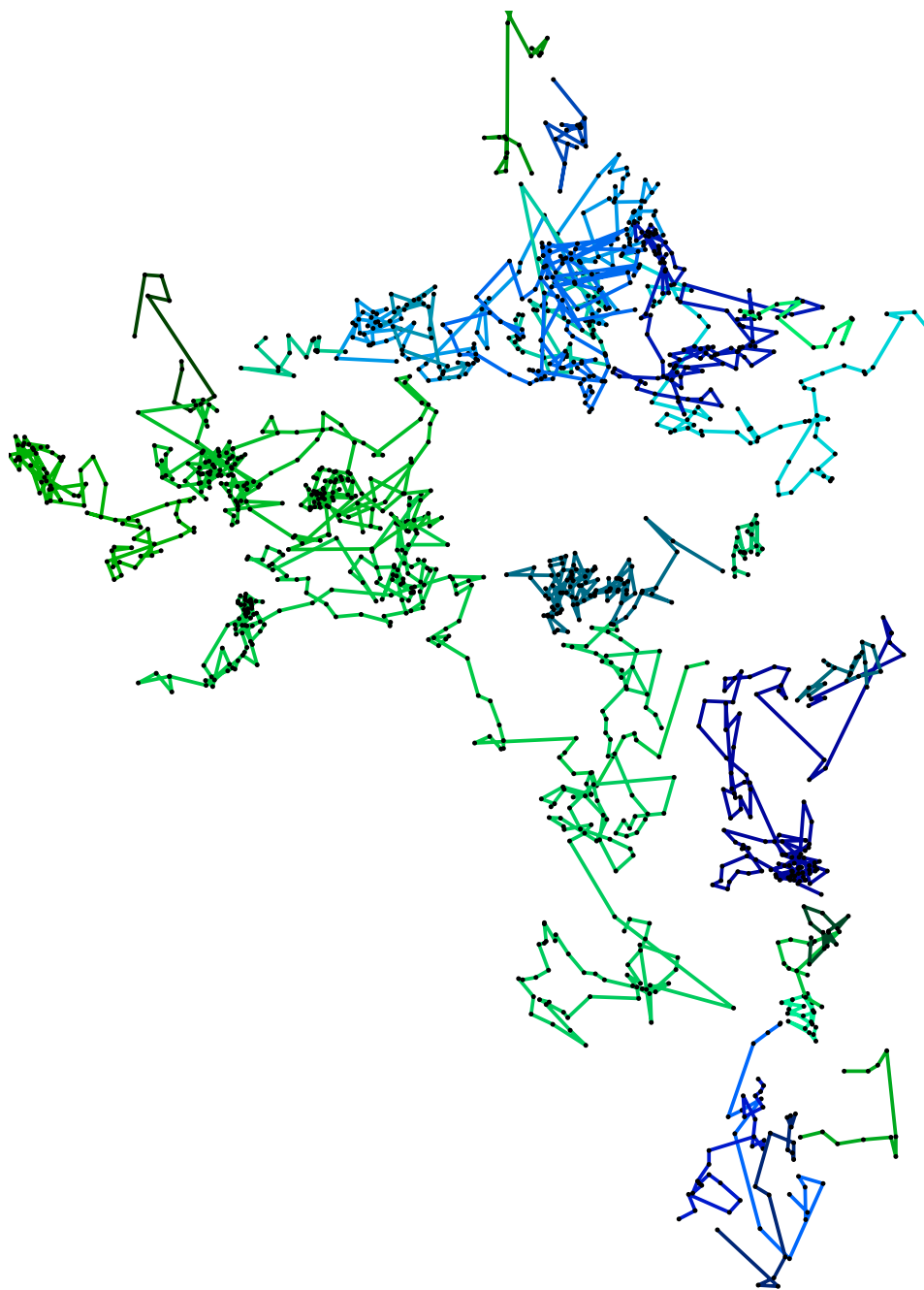
Figure 1.3: Several receptor trajectories derived from Single Particle Track-ing. We estimate the diffusion coefficients of monomer and dimer receptors from videos of these images.

involves tagging the EGF receptor with a fluorescent protein and taking pictures at fixed time steps of the fluorescent protein move across the cellular membrane. The caveat is determining where the particle is located since the wavelength of green light (or any other wavelength of visible light) is much larger than the protein. One can assume where the particle is by applying a normal distribution to the intensity of the green light. We can infer when a receptor is a monomer and a dimer through this method as well as the diffusion constant for the system. These observations will be useful when developing a spatial model for EGF; the first step towards the goal of a Well-Mixed model.

1.3 Mathematical Foundations of Modeling and Simulation

1.3.1 Chemical Reaction Networks

Chemical Reaction Networks describe any chemical reaction under certain situations. The main assumptions are that; the chemicals involved can be well represented by concentrations (the concentration of species S_i is denoted by $[S_i]$), which is an amount of substance in a given space (and as such are all non-negative) with base unit of concentration being *molar*, and the space and distribution of particles is uniform. Mentioning concentration, the molar is $1M = \frac{1mol}{1litre}$. We use CRNs to describe the state of cells which is a natural framework for the quantative description of the state and processes of a cell as it is basically a chemical factory. We can track most any change in the state of the cell using this framework such as a molecule of receptor migrating from the interior of the cell to the membrane or even kinase activation of that receptor[25].

Let us begin with an explanation of terms. First, a chemical species represents a type of molecule for example methane CH_4 or oxygen O_2. A reaction takes one or more chemical species and turns them into other chemical substance(s) such as $CH_4 + 2 \cdot O_2 \to CO_2 + 2 \cdot H_2O$. We should note that this is not a reversible reaction and is denoted by $\to$ whereas a reversible reaction would be denoted by $\rightleftharpoons$. We should also note that the above reaction takes place between one molecule of methane and 2 molecules of oxygen and generally the amounts of substance involved in the reactions are not the same as the number of grams of each substance reacting. In this example

assuming only the standard isotopes of each element, $16g$ of methane would react with $64g$ of oxygen (i.e. $1mol$ of methane to $2mol$ of oxygen) and not $1g$ of methane with $2g$ of oxygen.

CRNs are defined by the set of species they describe, $S_1, S_2, ..., S_n$, the reactions possible, $R_1, R_2, ...R_m$. The reaction R_j is subject to a chemical equation with a rate ϕ_j:

$$\sum_{i=1}^{n} \alpha_{ij}S_i \rightarrow \sum_{i=1}^{n} \beta_{ij}S_i; \phi_j = f_j([S_1], [S_2], ..., [S_n]) \tag{1.1}$$
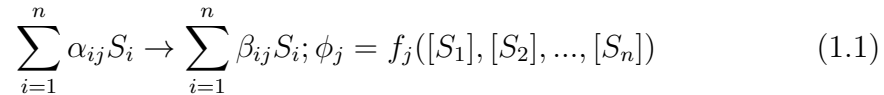
It is also clearly defined by the stoichiometric coefficients α_{ij}, β_{ij} along with the rate function $f_j(.)$ [25].

We generally gather the concentrations species at any time t into a vector $X = ([S_1], [S_2], ..., [S_n])^T$. If we look at the stoichiometric coefficients $\gamma_{ij} = \beta_{ij} - \alpha_{ij}$, they form a $m \times n$ matrix, $\Gamma = \gamma_{ij}$ [25]. If we also write out the fluxes as a column vector $\phi = (\phi_1, \phi_2, ..., \phi_m)^T$, we can write the equations of motion in the compact form:

$$\frac{dX}{dt} = \Gamma \cdot \phi(X) \tag{1.2}$$

Although several types of rate laws can be used the simplest is Mass Action Rate Laws. Assuming mass action, the rate for R_j becomes:

$$\phi_j = k_j \prod_{i=1}^{n} [S_i]^{\alpha_{ij}} \tag{1.3}$$

The rate constant k_j relates the molecular properties with the continuum description [25].

Numerical Solutions of ODE Models

Analytical solutions to many CRNs cannot be found in a closed form. This difficulty is due to several factors for example the number of variables involved or the fact that many of the equations are non-linear. Generally the point of a model is to produce typical behaviors of the system in question and these systems are usually initial value problems (IVPs). Due to these facts, more often than not the solution to a CRN is arrived using numerical methods. We begin discussing the mathematics used in this thesis with the Euler Method of

simulation. It is a simple and straight forward approach to solving differential equations assuming that the equations are both continuous and deterministic (we will discuss later why this is not accurate for chemical reactions except in special cases). Again the method is straight forward, and requires knowing only the n differential equations that describe the chemical reactions (this method can be applied to any differential equations, but I refer to chemical reactions because those are the only relevant equations to this thesis). We are concerned with systems of first-order ordinary differential equations (ODEs) that represent the time evolution of a set of state variables $\{X_1, \cdots, X_n\}$. We assume the equations of motion can be written in the explicit form

$$\frac{dX_i}{dt} = f_i(X_1, ..., X_n) \tag{1.4}$$

The caveat to the Euler Simulation is to pick a correct time step.

Any elementary numerical methods text should have a discussion of the Euler Algorithm, but I will explain it here for clarity.

The idea is to approximate the time derivatives from the definition

$$F'(t) = \lim_{\Delta t \to 0} \frac{F(t + \Delta t) - F(t)}{\Delta t} \quad \Rightarrow \quad F(t + \Delta t) - F(t) \approx \Delta t \cdot F'(t) \tag{1.5}$$

The above is a special case of using the Taylor series to approximate $F(t+\Delta t)$ using the value of the function and its derivatives at t:

$$F(t + \Delta t) = \sum_{k=0}^{m} \frac{(\Delta t)^n}{k!} F^{(k)}(t) + \mathcal{O}((\Delta t)^{m+1}) \tag{1.6}$$

Here, $\mathcal{O}(\Delta t^{m+1})$ represents the remainder term, which is proportional to Δt^{m+1} (i.e. it is of the form $R_m = F^{(m+1)}(\xi)\Delta t^{m+1}/(m + 1)!$, where $\xi \in [t, t + \Delta t]$; hence in the limit $\lim_{\Delta t \to 0}$ it goes to zero like Δt^{m+1}). The Euler method corresponds to a first order approximation:

$$F(t + \Delta t) = F(t) + F'(t)\Delta t + \mathcal{O}(\Delta t^2) \quad . \tag{1.7}$$

where the remainder term is proportional to Δt^2 (this term will go to 0 much more quickly than Δt showing that it converges as Δt goes to 0 also giving validity to our approximation method and meaning that the time step must be sufficiently small).

Once a sufficiently small time step has been picked the n differential equations are multiplied by the time step. To see this consider the following:

$$\Delta X_i(t_0) = f_i(X_1(t_0), ..., X_n(t_0)) \cdot \Delta t$$
$$X_i(t_0 + \Delta t) = X_i(t_0) + \Delta X_i \tag{1.8}$$

This is one of the simplest, if not the simplest methods for simulation. We only need to repeat the process until we reach the end of our simulation time.

1.3.2 Stochastic Model / Simulations

A mathematical model is a mathematical system that emulates the system of interest. Mathematical modeling has been utilized in the field of Physics for hundreds of years. Models are useful because the help with our understanding of the real system (i.e. how it works) and can make predictions on how the system will behave. A simulation is a prediction of a model with regards to most often a specific experiment. When investigating an ODE model, the solution of the simulation (arrived at either numerically or analytically) predicts the exact future behavior of the system as described by the model. This exactness is not the case with a stochastic simulation however. Instead for stochastic models the future behavior is a distribution and simulations provide one possible version of the future.

Poisson Processes

We now switch to a discussion of Poisson Processes, which are necessary because the ODE approach implies a continuum of uniformly distributed molecules, but molecules are particles and as such they occur in discrete quantities, so this approach gives a better approximation for the CRNs of cellular processes. However the ODE approach becomes more "correct" as the number of molecules increase (we should note that Avogadro's number is the number of molecules in one mole of substance $N_{Avogadro} \approx 6 \cdot 10^{23}$). Yet the number of copies of a given chemical substance within a cell can be very small (in a human cell a chemical with a concentration of $100nM$ has approximately 60 copies). Considering receptors, we must also account for the spatial structure of the cellular membrane and we have seen from the TEM images and SPT data that the receptors cluster, because of this we also

cannot assume uniform distribution of molecules further removing ourselves from the continuum and uniformity requirements [25].

The easiest example of a Poisson Process to understand is radioactive decay. Let us assume we have an element A (with a number of particles N_A) which can undergo radioactive decay given by $A \rightarrow \emptyset$, with a rate k in inverse time. If we look at it through the eyes of ODEs and Mass Action Rate Laws we arrive at the differential equation $\frac{dA}{dt} = -k \cdot A$. An elementary Calculus book would give the solution as $N_A(t) = N_A \cdot e^{-k \cdot t}$, but this solution is only valid when dealing with a continuum.

Our first concern is when a single particle of A will decay. The probability that A will decay in a given time interval $(t, t + \Delta t)$ is proportional to Δt.

$$\lim_{\Delta t \to 0} \frac{\Delta p_A(t, t + \Delta t)}{\Delta t} = k \rightarrow \Delta p_A(t + \Delta t) \approx k \cdot \Delta t \tag{1.9}$$

If we consider N_A to be larger than 1 but not large enough to justify using the continuum approach, then the number of particles of A that remain $n_A(t)$ over $(t, \Delta t)$ is approximated by:

$$n_A(t) - n_A(t + \Delta t) \approx \Delta p_A(t, t + \Delta t) \cdot n_A(t) + \mathcal{O}(\Delta t^2) \tag{1.10}$$

The above approximate equalities neglect differences on the order of $\frac{1}{\sqrt{\Delta n_A}}$. Assuming $n_A(0) = N_A$, molecule numbers comparable to $N_{Avogadro}$ and a time step $\Delta t \ll \frac{1}{k}$ we can treat n_A as continuous [25]. The above imply the following equation and solution:

$$\lim_{\Delta t \to 0} \frac{n_A(t + \Delta t) - n_A(t)}{\Delta t} = \frac{dn_A}{dt} = -k \cdot n_A(t) \tag{1.11}$$

$$n_A(t) = N_A \cdot e^{-k \cdot t} \tag{1.12}$$

This agrees with our earlier solution and demonstrates the connection with the continuum approach.

Combining Poisson Processes

Poisson Processes have an interesting feature with regards to combining processes. Let us assume we have two independent Poisson Processes with rates k_1 and k_2, and we are only concerned with the first time either of them occur. Starting at $t = 0$, the total rate is $k_T = k_1 + k_2$ and the probability that event 1 occurs first is $p_1 = \frac{k_1}{k_T}$. If we consider the previous section, we

verify that the probability to have $n_A(t)$ at time t is given by the probability $g_{A \to \emptyset} = n_A \cdot k$.

The time any A would decay is given by the PDF:

$$f(\tau) = g_{A \to \emptyset} \cdot e^{-g_{A \to \emptyset} \tau} \tag{1.13}$$

This principle of Poisson processes can also be applied to when we have several different types of reactions [25]. In general, we can combine any number of Poisson processes using an inductive argument (i.e. if we have n processes with rates $k_1, k_2, \cdots, k_n$ and denoting the total again as k_T the probably that event i occurs first is $\frac{k_i}{k_T}$). We will use this property of Poisson processes for our simulations.

The Gillespie Stochastic Simulation Algorithm (SSA)

"The differential reaction-rate equations approach to chemical kinetics cannot be denied, we should not lose sight of the fact that the physical basis for this approach leaves something to be desired." [22] The approach assumes chemical reactions to be continuous and deterministic, but molecular changes are in integer amounts [22]. Because of this the Euler Simulation is only correct for large numbers of molecules where we can assume a continuous amount of substance. For small numbers a different method is required. We will use the Gillespie Algorithm (more precisely the Next Event Method which will be discussed in detail later) to simulate the time evolution of chemical processes. The Gillespie Algorithm "take explicit account of the fact that the time evolution of a spatially homogeneous chemical system is a discrete, stochastic process instead of a continuous, deterministic process." [22]

Chemical reactions, in general, occur during collisions between molecules [22] the exception being unimolecular reaction such as radioactive decay. At thermal equilibrium, the collisions and subsequent reactions occur randomly [22]. One advantage to the Gillespie Algorithm is that it is equivalent to the Chemical Master Equation approach [22], but this author finds it much easier use the Gillespie approach.

The algorithm is fairly easy to implement. Assume one has n distinct reactions, with propensities call them $\Phi_1 \dots \Phi_n$ that are the rates for Poisson processes (in our previous example $g_{A \to \emptyset}$). We then generate a random number for each of the propensities of the reactions Φ_i (given by the PDF

13

of reaction i). This random number is the time, t_i that reaction i will occur. We allow the reaction m to occur if t_m is the smallest time among times generated. Because each reaction rate depends on the number of each molecule present, we then update the amount of substances due to reaction m and generate new random numbers for the reactions. We then continue this process until we have reached the time limit for our simulation [22].

Spelling it out in more detail, we obtain a differential equation $n'_A(t) = -kn_A(t)$ and should note it has the same form as mass action if we replace n_A with $[A]$. Because concentration depends directly on the number of molecules, we can approximate the ratios $\frac{[A]'}{[A]} \approx \frac{n'_A}{n_A}$ and the propensity per particle of A is:

$$\frac{[A]'}{[A]} = \lim_{V \to \infty} \frac{g_{A \to \emptyset}}{n_A} = -k \tag{1.14}$$

The Gillespie Algorithm generates one possible history of the chemical reactions. This history will not be the same from one running of the simulation to the next because of the random nature of stochastic simulations. This method has advantages over the deterministic approach, but is computationally expensive if the number of reactions is large enough. The generation of many random numbers is taxing on the computer. This is why we will consider using the Next Event Method.

Next Event Method

The Next Event Method is an improvement upon the Gillespie Algorithm using the property of combining Poisson processes. Again assume we have n reactions, instead of generating n random numbers as with the Gillespie Algorithm, the Next Event Method only requires two random numbers per update. Since we have propensities of n reactions Φ_1 ... Φ_n, the Next Event Method looks at the total probability of any event occurring, Φ_T, and generates a time based on that total. This t is one random number necessary for the algorithm. The next random number necessary picks the reaction that occurs by dividing the propensity of the reaction Φ_i by propensity of Φ_{total} and assigning the probability $\frac{\Phi_i}{\Phi_{total}} = P_i$ to reaction Φ_i. By listing the probabilities in a vector $(P_1, P_1 + P_2, ..., 1)$ then generating one random number between 0 and 1, we pick the reaction that occurs.

This method is computationally less expensive than the standard Gillespie Algorithm for large numbers of reactions. Because of this fact all stochastic

chemical reaction simulations by this author are done using this algorithm.

1.3.3 Diffusion, Smoldyn and Spatial Models

Brownian Motion (Diffusion)

Brownian motion is a random walk, meaning that given the position of a particle (x_0, y_0) at $t = 0$ we cannot know the position at any future time, but only know the probability of the location. Consider the position vector of a particle, $r(t) = (x(t), y(t))$. The two dimensional motion of particle is considered Brownian if the components of the displacement Δx and Δy of the displacement vector $\Delta r = r(t + \Delta t) - r(t)$ over the time interval $(t, t + \Delta t)$ are distributed according to the following PDF [25]:

$$f(\Delta x, \Delta y; \Delta t) = \frac{1}{4\pi D \Delta t} e^{\frac{\Delta x^2 + \Delta y^2}{4D\Delta t}} \tag{1.15}$$

This means that the variables x and y are distributed independently with variance $\sigma^2 = 2D \cdot \Delta t$ where D represents the diffusion constant (a measure of the particles mobility) [25]. Since Brownian motion is random, even knowing the position (x_0, y_0) at time $t = 0$ we can only know the probably of the position of the particle at a later time. The localization probability density agrees with the PDF (1.15) from above with $\Delta x = x - x_0$, $\Delta y = y - y_0$ and $\Delta t = t$.

$$p(x, y; t) = \frac{1}{4\pi Dt} e^{-\frac{(x-x_0)^2 + (y-y_0)^2}{4Dt}}$$

We should note that this is normalized for all $t > 0$ (i.e. $\iint p(x, y; t) dx dy = 1$). The evolution in time is consistent with the PDF(1.15). If we let $\Delta t = s$ for simplicity, the localization density at time $t = t_0 + s$ is given by the equation:

$$p(x, y : t_0 + s) = \iint p(x', y'; t_0) f(x - x', y - y'; s) dx' dy'$$

If we ignore the issue of overlapping particles, we can describe n particles undergoing Brownian motion with the sum of the individual localization probability density functions $p^j(x, y; t)$ and the joint function in normalized to the number of particles n.

$$p(x, y; t) \equiv \sum_{j=1}^{n} p^j(x, y; t); \iint p(x, y; t) dx dy = n \tag{1.16}$$

The equation defined in (1.16) also verify the diffusion equation:

$$\frac{\partial p}{\partial t} = D\left(\frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2}\right) \tag{1.17}$$

Brownian motion simulations

To simulate simple Brownian motion on a computer, we need to keep track of the position of each particle (i.e. $\{(x_j, y_j, s_j)\}_{j=1\cdots N_p}$ where N_p is the total number of particles). Given the position of each particle $\{r_j = (x_j, y_j)\}_{j=1\cdots N_p}$ at time t the coordinates at time $t + \Delta t$ are obtained by generating random numbers according to (1.15). A general implementation would make use of a fixed time step Δt depending on the other elements of the simulation.

Now consider simulating Brownian motion of receptors on a cellular membrane. We must take into account the computational limitations as the cellular membrane is too large to simulate the entire area because of the necessarily short time steps required by other considerations of the simulation. Due to this, we will focus on smaller patches of the membrane (say a box $[0, B_x] \times [0, B_y]$ with a proposed update of (x^*, y^*)) and apply reflecting or periodic boundary conditions.

If we apply reflecting boundary conditions, if $x^* < 0$ we update with $x^* \to -x^*$ and if $x^* > B_x$ we update with $x^* \to (B_x - x^*)$. We'll use a similar procedure for y^*. If instead we consider periodic boundary conditions, if $x^* > B_x$ we update with $x^* \to (x^* - B_x)$ and for $x^* < 0$ we update with $x^* \to (x^* + B_x)$ and again a similar procedure is used for y^*. However, both procedures have attached artifacts that have no relation to reality such as strange geometries. When we consider particles that interact, the reflecting boundary conditions can also cause particles to pile up about the boundaries so instead we will use periodic boundary conditions [25].

Now we need to concern ourselves with combining reactions with diffusion. First consider first order reactions (like radioactive decay) we can treat the diffusion as independent of the reaction. We now not only have to keep track of the position of each particle but also its chemical state. We will chose the time a reaction occurs similarly to a well-mixed simulation by using the PDF of the reaction (1.13) (call it τ_{React}), but now we must also chose the individual particle which reacts. This can be done by generating a random number to pick the particle reacting. We do this by generating a random number $r\epsilon[0, 1]$ and partitioning the particles as $\{\frac{1}{n_A}, \frac{2}{n_A}, \cdots \frac{n_A-1}{n_A}\}$.

The random number r falls into any given subinterval $[\frac{j-1}{n_A}, \frac{j}{n_A}]$ is the same for all j and the resulting j determines the transitioning particle [25]. Since the time of reaction is independent of diffusion and the particle number does not change in first order reactions where $A \rightarrow B$, we can apply the previous discussion of Brownian motion (1.15) to all particles from $(t, t + \tau_{React})$.

Things become more complicated when we consider second order reactions. Consider reaction $A+B \rightarrow \cdots$. For the reaction to occur the molecules must be in close proximity which we can judge based on the distance between the position vectors of the particles $r_j = (x_k, y_k)$ and knowledge of the size of the molecules. If we consider two molecules with position vectors r_A, r_B then we can tell if they are close enough to react based on the distance:

$$|r_A - r_B| \leq d_{AB}^{coll} \tag{1.18}$$

Once we know that two molecules have collided we need to determine if those molecules react or bounce off of each other. If the molecules do not react (an elastic collision), we have to resolve the fact that the molecules may now overlap which is physically impossible. One approach is similar to the solution of reflective boundaries, and another is to leave the colliding particles in their original position. The main issue now is to determine if the particles that have collided react. We can do this one of two ways: the first would be to generate a random number and the other is to use a binding radius that is smaller than the molecules. Another consideration is to ensure that the time step is sufficiently small enough to avoid missing collisions.

Smoldyn

To make use of the SPT data, we need to construct a spatial model for the EGF system from the inferred diffusion constant, dimerization rate and dissociation rate we discovered from the STP data. We use an algorithm based on Smoldyn which simulate both Brownian motion as described above and chemical reactions. In Smoldyn, a time step Δt, diffusion coefficient D_0, binding radius (BR) and unbinding radius (UBR) are needed [3]. The diffusion coefficient is used to simulate the motion of the receptors along the cellular membrane as described in the previous section. The binding radius takes into account the likelihood of a binding event given a collision and is generally smaller than the physical size of the reacting molecules (this prevents us from needing to generate a random number to determine if colliding molecules react making the simulation run faster). The binding radius can be

determined by comparing the SPT data to a Smoldyn simulation and adjusting the binding radius until the simulation matches the data. The unbinding radius is taken to be 5 times the binding radius as suggested by Andrews and Bray to prevent rebindings [3]. The time step Δt needs to be chosen to be small enough to avoid missing collisions between molecules.

Multi-parameter Nonlinear Optimization

The objective is to derive kinetic coefficients (i.e. k_{on} and k_{off}) for a well-mixed model of EGF using the spatial simulation data. First I have found an analytical solution for the EGF system and then compare the distance between the analytical solution and the spatial simulation at several of points. There are several methods for calculating extrema however the basic calculus method using the first and second derivative cannot be used, but instead other options do exist. The Metropolis-Hastings Algorithm allows for directed but random choices for our kinetic parameters. This algorithm will randomly walk the parameters around the space of possible values until either a max or min is achieved (in our case a minimum value of the distance from the spatial values at given time points to the solution using the randomly chosen parameter values). This extreme value will then serve as our parameters for the Well-Mixed EGF model.

Chapter 2

Coarse grained model of VEGF signal initiation

This chapter of my thesis introduces a coarse grained approach to incorporate the attractive (trapping) regions for receptors found on the cell membrane through TEM images and SPT data with an existing Chemical Reaction Network (CRN) specifically applied Vascular Endothelial Growth Factor (VEGF). This approach can be applied to any well-mixed model of species occurring in two dimensions with regions that collect the chemicals in question. The effects of these domains on cell signaling needs to be understood as any change to the concentration of signaling complexes could drastically alter the cell's response to said signal.

2.1   Biology of VEGF

In the specific system we are studying, VEGF, the receptors are monovalent, meaning that they have one ligand binding site; however, the ligand is bivalent, which allows it to bind to two receptors[6]. The VEGF receptors (VEGFR) can also dimerize and the two receptors may dimerize even in the absence of the ligand[6, 21]. In this case, receptors move in conjunction, but are not activated (phosphorylated in this case) and, as such, are not transmitting signals into the cell [21](See Figure 2.1).

VEGFR belong to the tyrosine kinase receptor family, with extracellular ligand binding and receptor dimerizing domains and intracellular kinase and regulatory domains. Ligand binding to the receptor is followed by receptor
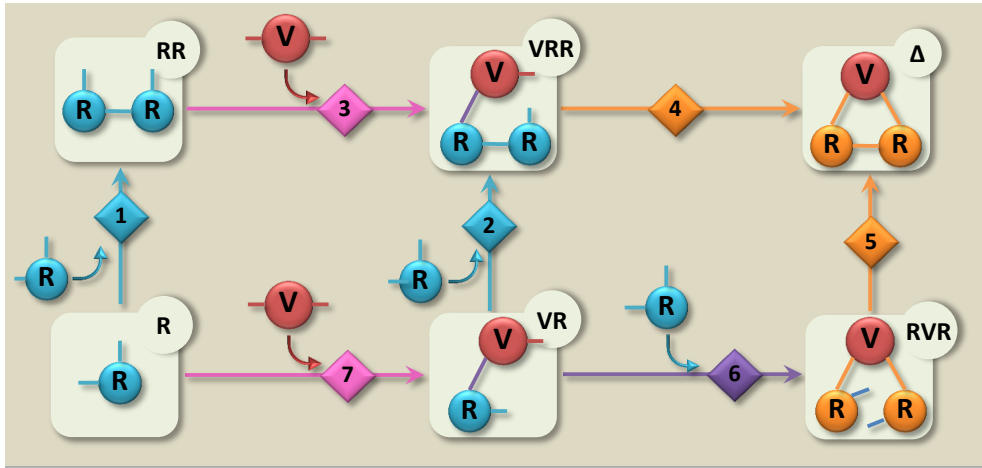
Figure 2.1: Reactions and species for VEGF with a single receptor species (based on [21]). This representation shows all species and reactions involved in the local reaction network. Note that both species on the right (Δ as well as RVR) are considered signaling.

dimerization, activation of the kinase domain and transphosphorylation of specific tyrosine residues on the intracellular domain of the receptor [43]. The phosphorylated tyrosine residues then serve as docking sites for other proteins including the Src Homology 2 (SH2 domain) containing proteins. [21, 6]. These SH-2 domain containing proteins continue binding other proteins until the signal is finally transmitted to the cell.

VEGF signaling is a major health concern because of the connection between VEGF and angiogenesis (formation of new blood vessels from existing vascular system) [56]. In normal and healthy cellular function, for example, VEGF is secreted after serious muscle strain or damage [7]. Another normal cellular function of VEGF is in organ development [56]. The effects of exercise on angiogenesis via VEGF have been seen in mice with a significant difference in blood vessel growth [33]. VEGF begins angiogenesis to both repair damaged muscle tissue and create new vasculature to provide more nutrients and oxygen so that the muscle tissue is better able to cope the next time such exercise is performed. Other normal and healthy functions of VEGF include: summoning blood vessels in wound repair[35] and embryonic blood vessel development and creation [59]. An embryo would not survive past day 11 without one VEGFA allele [52]. However it does not appear to be necessary in maintaining vasculature once created [59].

Concentration and gradient of VEGF signal have a role in angiogenesis [56]. "The spatial distribution of VEGF is a key regulator of angiogenesis" [56]. Computational models have shown that the gradient of VEGF cannot be explained by heparan sulfate proteoglycan (HSPG) bindings alone [56]. VEGF has several isoforms which lead to different types of blood vessel formation [56]. Heparin-binding VEGF isoforms create narrow branching blood vessels, whereas $VEGF_{120}$ produces leaky vessels [56].

Despite the necessity of VEGF in normal cellular function, angiogenesis also has a role in cancerous cellular division and proliferation [14]. Generally, tumors overexpress $VEGF_{121}$ [56]. In this context, VEGF contributes to tumor growth beyond a small size limit by supplying the oxygen required for continued cellular division and VEGF is a key ligand in this process. This new vascular network not only provides the required oxygen and nutrients for growth[9], but can also provide a path for the tumor to metastasize and spread into other organs and tissues[6]. These new tumors can prove fatal and, as such, are a major health concern. Expression of $VEGF_{164}$ shows the most rapid tumor growth [56]. "VEGF-A also promotes a wide range of functions, both in vitro and in vivo, all autocrine function on tumor cells, including adhesion, survival, migration and invasion." [47]

Pharmaceutical research has been aimed at preventing VEGF signaling in order to inhibit or at least slow the process of angiogenesis [1]. Avastin has been FDA approved for the treatment of some cancers [52]. Some preclinical mouse models have shown promise in this area by decreasing the mass of tumor [42]. "Clinically, control of vessel morphogenesis through the VEGF family would be of use in a wide range of diseases." [56]

VEGF-A plays a role in age-related macular degeneration (AMD) which is the leading cause of blindness in the elderly [31]. However, the response in these cases to anti-VEGF drugs in not consistent [31] VEGF-A is also plays a role in diabetic nephropathy(DN) [55] DN is "the leading cause of end-stage renal disease worldwide [55]. A better understanding of the VEGF system should allow for a better understanding of how effective these drugs need to be in order to hinder tumor related angiogenesis as well as other diseases.

Looking only at VEGF Receptor-1 (VEGFR1) and the ligand VEGF-A gives rise to a seven species and fourteen reaction system (see Figure 1). This is complicated by the existence of several distinct variants of the VEGF receptor, which overlap on the type of VEGF ligand they can bind, as well as several variants of the VEGF ligand. If we consider only one additional receptor type (VEGFR2, which can also bind with VEGFA), the number of

reactions possible grows to almost fifty[21], which gives rise to combinatorial explosion if we consider all possible receptor and ligand types. Because of this, we will restrict ourselves to the homodimer case of one type of VEGF ligand and receptor.

2.2   Model Building

As mentioned earlier, VEGF receptors reside on the cell membrane and TEM images and SPT data show a non-homogeneous spatial distribution of those receptors. Due to this clustering of the receptors, I develop a transfer network of domains with the connection information of the network compactly contained in matrix which is referred to in this thesis as the adjacency matrix.

To simulate the VEGF chemical reaction network, the stoichiometric coefficients of both the inputs, and outputs contained in matrices call them α and β respectively, must be obtained from information about the local reaction network and the transfer network. The construction will be done in pieces with the first being the total local behavior followed by the movement behavior. Finally, the two pieces will be combined.

In order to facilitate analytics, we will also break the total VEGF system into two simpler components and consider them individually. Firstly, we will consider the Ligand Induced Dimerization (LID) only system which allows for dimerization only in the presence of the VEGF ligand. Secondly, we will also consider the VEGF system in the absence of ligand which allows only Dynamic Pre-Dimerization (DPD). The DPD system is similar to a completely liganded EGF system. The analytics of the LID system are included in the text; the DPD system will be omitted as it is basically the fully liganded EGF system.

2.2.1   Defining the problem

Local reaction network.   Consider a *generic chemical reaction network* (CRN) with N_S molecular species $\mathcal{S} \equiv \{\mathsf{S}_1, \cdots, \mathsf{S}_{N_S}\}$; we will call it *local* to distinguish from the composed system. The concentrations form the state vector,

$$\mathbf{X}_{\mathrm{loc}} = (X_1, X_2, \cdots, X_{N_S})^T \quad , \quad \text{where} \quad X_k \equiv [\mathsf{S}_k] \quad . \tag{2.1}$$

The N_R (local) reactions are defined by two sets of stoichiometric coefficients, $\{\alpha_{ij}, \beta_{ij}\}_{1 \le i \le N_R, 1 \le j \le N_S}$, and we assume that the reactions follow mass action laws with rate constants k_i:

$$\sum_{j=1}^{N_S} \alpha_{ij} \mathsf{S}_j \xrightarrow{k_i} \sum_{j=1}^{N_S} \beta_{ij} \mathsf{S}_j \quad ; \quad \phi_i = k_i \prod_{j=1}^{N_S} X_j^{\alpha_{ij}} \quad , \ (\forall) \ 1 \le i \le N_R \ . \quad (2.2)$$
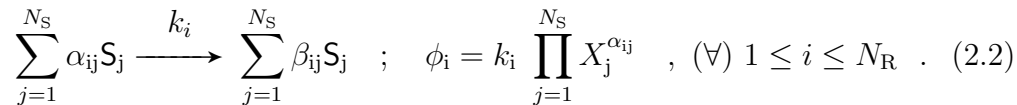
It is useful to collect the rate constants into a vector $\mathbf{K}_{\mathrm{loc}} \equiv (k_1, \cdots, k_{N_R})^T$, and to define net stoichiometry coefficients $\gamma_{ij} \equiv \beta_{ij} - \alpha_{ij}$. The equations of motion for the local CRN can be written in matrix form:

$$\frac{d\mathbf{X}_{\mathrm{loc}}}{dt} \ = \ \Gamma_{\mathrm{loc}} \cdot \Phi_{\mathrm{loc}}(\mathbf{X}_{\mathrm{loc}}) \quad , \quad (2.3)$$

where $\Gamma_{\mathrm{loc}} \equiv \{\gamma_{ij}\}_{1 \le i \le N_S, 1 \le j \le N_R}$ is the $N_S \times N_R$ local stoichiometric matrix and $\Phi_{\mathrm{loc}} \equiv (\phi_1, \cdots, \phi_{N_R})^T$ is the $N_R \times 1$ vector of local reaction fluxes.

Network of locations.   We assume that the local CRN defined above is realized *in N_L different locations*, $\mathcal{L} \equiv \{\mathsf{L}_1, \cdots, \mathsf{L}_{N_L}\}$. Each copy of the CRN has the same set of reactions and (physical) rate constants,[1] with generally different amounts of each species. In addition to the local, molecular transformations, the N_L locations may exchange substances, depending on their proximity and physical characteristics. We will assume that substance transfer is limited to a subset of the possible (oriented) pairs of locations.

The possible connections are summarized in a weighted adjacency matrix, $\mathbf{A} = \{a_{l,m}\}_{1 \le l, m \le N_L}$, that defines a connectivity graph. All nonzero entries of $\mathbf{A}$ are positive, and a_{lm} is a measure of the transfer rate *from* location l *to* location m. The connections and their strengths define the possibility and relative ease of transfer between respective locations. We describe the actual transfer processes as mass-action reactions, involving the respective species in the origin and destination location. The transfer rate constant is determined by the connection strength a_{lm} and a mobility factor μ_j that characterizes each molecular species. For example, transfer of substance S_j from location l to location m has the rate

$$\psi_j^{(l \to m)} \ = \ a_{lm} \cdot \mu_j \cdot [\mathsf{S}_j^{(l)}] \quad , \quad (2.4)$$

---

[1]Locations will typically differ in terms of their physical extent, and this may impact the *effective* reaction rate constants, even if the *physical* rate constants are the same. Of course, the physical rate constants for corresponding reactions in different locations may also differ, but for the sake of a clearer presentation we will assume them to be identical.

where $[\mathsf{S}_{\mathrm{j}}^{(1)}]$ denotes the concentration of substance S_{j} in (the source) location l.

We do not make any assumptions regarding symmetry: links may be asymmetric, or even uni-directional. Since we generally expect $\mathbf{A}$ to be sparse, it is useful to characterize its nonzero entries; let $\{(\rho_{\mathrm{c}}, \sigma_{\mathrm{c}})\}_{1 \leq c \leq N_{\mathrm{C}}}$ be the list of nonzero entries of A, and denote by ω_{c} the corresponding connection strengths,

$$a_{\mathrm{ij}} = \sum_{c=1}^{N_{\mathrm{C}}} \delta_{\mathrm{i},\rho_{\mathrm{c}}} \delta_{\mathrm{j},\sigma_{\mathrm{c}}} \omega_{\mathrm{c}} \quad ; \quad {}^{(\forall) 1 \leq l \leq N_{\mathrm{L}}, \ 1 \leq m \leq N_{\mathrm{L}}} \quad . \tag{2.5}$$

where for the N_{C} nonzero entries in the matrix, the column index σ_{c} gives the exiting location and the row index ρ_{c} yields the entry location; we used the Kronecker delta: $\delta_{\mathrm{jl}} = 1$ if $j = l$ and $\delta_{\mathrm{jl}} = 0$ if $j \neq l$.

2.2.2 Putting things together

Let us now turn to the *joint system* formed by all substances at all locations. This system will have $N_{\mathrm{TS}} = N_{\mathrm{S}} \cdot N_{\mathrm{L}}$ molecular species, corresponding to the species-location pairs of the Cartesian product

$$\mathcal{T} \equiv \mathcal{L} \times \mathcal{S} = \{\mathsf{L}_1\mathsf{S}_1, \mathsf{L}_1\mathsf{S}_2, \cdots, \mathsf{L}_2\mathsf{S}_1, \cdots, \mathsf{L}_{N_{\mathrm{L}}}\mathsf{S}_{N_{\mathrm{S}}}\} \tag{2.6}$$

The entries of the full state vector correspond to the N_{TS} concentrations, rearranged into a single column vector:

$$\begin{aligned} \mathbf{Y} &= ([\mathsf{L}_1\mathsf{S}_1], [\mathsf{L}_1\mathsf{S}_2], \cdots, [\mathsf{L}_{\mathrm{m}}\mathsf{S}_{\mathrm{j}}], \cdots, [\mathsf{L}_{N_{\mathrm{L}}}\mathsf{S}_{N_{\mathrm{S}}}])^T = \\ &= \left(Y_1, Y_2, \cdots, Y_{(\mathrm{m}-1)\mathrm{N}_{\mathrm{S}}+\mathrm{j}}, \cdots, Y_{\mathrm{N}_{\mathrm{TS}}}\right)^T \end{aligned} \tag{2.7}$$

Our goal is to summarize the molecular reactions as well as the transfer reactions between the elements of $\mathcal{T}$ as mass-action reactions, characterized by stoichiometry matrices and rate constants as in Eq. (2.2), leading to equations of motion similar to Eq. (2.3).

Molecular processes: The $N_\mathrm{R} \cdot N_\mathrm{L}$ *molecular transformations* will have the same form as the local reaction set,

$$\sum_{j=1}^{N_\mathrm{S}} \alpha_{ij}(\mathsf{L_m S_j}) \xrightarrow{\quad k_i^{(\mathrm{m})} \quad} \sum_{j=1}^{N_\mathrm{S}} \beta_{ij}(\mathsf{L_m S_j}) \quad ;$$

$$\phi_\mathsf{i}^{(\mathrm{m})} = k_\mathsf{i}^{(\mathrm{m})} \prod_{j=1}^{N_\mathrm{S}} Y_{(\mathrm{m}-1)\mathrm{N_S}+\mathsf{j}}^{\alpha_{ij}} \quad , \ (\forall) \ 1 \leq i \leq N_\mathrm{R} \quad , \tag{2.8}$$

where $\phi_\mathsf{i}^{(\mathrm{m})}$ is the flux, and $k_\mathsf{i}^{(\mathrm{m})}$ is the mass-action rate constant of reaction i occurring in location m. We form the vector Φ_mol of molecular transformations and order the rate constants in a vector $\mathbf{K}_\mathrm{mol}$ following the same ordering convention as in Eqs. (2.6,2.7):

$$\Phi_\mathrm{mol} \ \equiv \ \left(\phi_1^{(1)}, \phi_2^{(1)}, \cdots, \phi_{\mathrm{N_R}}^{(1)}, \phi_1^{(2)}, \cdots, \phi_{\mathrm{N_R}}^{(\mathrm{N_L})} \right)^T$$

$$\mathbf{K}_\mathrm{mol} \ \equiv \ \left(k_1^{(1)}, k_2^{(1)}, \cdots, k_{\mathrm{N_R}}^{(\mathrm{N_L})} \right)^T \tag{2.9}$$

We will sometimes assume that the values of the local rate constants are independent of the location: $k_\mathsf{j}^{(\mathrm{m})} = k_\mathsf{j}$, $(\forall)$ $m = 1 \cdots N_\mathrm{L}$. In this case, it is useful to write the full (molecular) rate constant vector as

$$\mathbf{K}_\mathrm{mol} \ = \ \mathbb{1}_{\mathrm{N_L} \times 1} \otimes \mathbf{K}_\mathrm{loc} \quad , \tag{2.10}$$

where $\mathbb{1}_\mathrm{m \times n}$ is an $m \times n$ matrix whose entries are 1, and $\otimes$ denotes the Kronecker product of two matrices (explained in the appendix)

Irrespective of the rate constants, the left and right stoichiometry matrices can always be written as Kronecker products involving the $\alpha_\mathrm{loc} \equiv \{\alpha_{ij}\}$, $\beta_\mathrm{loc} \equiv \{\beta_{ij}\}$ matrices and the identity matrix $\mathbb{I}_{\mathrm{N_L} \times \mathrm{N_L}}$:

$$\alpha_\mathrm{mol} \ \equiv \ \mathbb{I}_{\mathrm{N_L} \times \mathrm{N_L}} \otimes \alpha_\mathrm{loc} \ \ ; \ \ \beta_\mathrm{mol} \ \equiv \ \mathbb{I}_{\mathrm{N_L} \times \mathrm{N_L}} \otimes \beta_\mathrm{loc} \tag{2.11}$$

Finally, defining the net stoichiometry matrix Γ_mol, the equations of motion corresponding to molecular transformations can be written as

$$\left(\frac{d\mathbf{Y}}{dt} \right)_\mathrm{mol} = \ \Gamma_\mathrm{mol} \cdot \Phi_\mathrm{mol}(\mathbf{Y}) \quad ; \ \Gamma_\mathrm{mol} \ \equiv \ \mathbb{I}_{\mathrm{N_L} \times \mathrm{N_L}} \otimes \Gamma_\mathrm{loc} \quad . \tag{2.12}$$

Transfer processes: We use the sparse matrix notation (2.5) for the nonzero entries of the adjacency matrix $A = \{a_{\mathrm{lm}}\}$. The transfer process corresponding to element $a_{\rho_c \sigma_c} = \omega_c$, taking substance S_{j} from location L_{ρ_c} to location L_{σ_c}, is described as follows:

$$(\mathsf{L}_{\rho_c}\mathsf{S}_{\mathrm{j}}) \xrightarrow{\kappa_{\mathrm{j}}^{(c)}} (\mathsf{L}_{\sigma_c}\mathsf{S}_{\mathrm{j}}) \quad ; \quad \psi_{\mathrm{j}}^{(c)} = \kappa_{\mathrm{j}}^{(c)}[\mathsf{L}_{\rho_c}\mathsf{S}_{\mathrm{j}}]$$

$$(\forall) 1 \leq j \leq N_{\mathrm{S}}, \ 1 \leq c \leq N_{\mathrm{C}} \ . \tag{2.13}$$

The transfer rate constant $\kappa_{\mathrm{j}}^{(c)}$ is determined by the strength of the location-to-location connection ω_c and the mobility μ_{j} of the substance,

$$\kappa_{\mathrm{j}}^c = \omega_c \cdot \mu_{\mathrm{j}} \quad , \quad (\forall) 1 \leq j \leq N_{\mathrm{S}}, \ 1 \leq c \leq N_{\mathrm{C}} \ . \tag{2.14}$$

We form the vector Ψ_{trans} from the fluxes $\psi_{\mathrm{j}}^{(c)}$, by enumerating the substances first, then the processes:

$$\Psi_{\mathrm{trans}} \equiv \left(\psi_1^{(1)}, \psi_2^{(1)}, \cdots, \psi_{N_{\mathrm{S}}}^{(1)}, \psi_1^{(2)}, \cdots, \psi_{N_{\mathrm{S}}}^{(N_{\mathrm{C}})} \right)^T \ . \tag{2.15}$$

The fluxes represent the $N_{\mathrm{C}} \cdot N_{\mathrm{S}}$ transfer processes that will be represented as an additional set of mass-action reactions in the joint system. The vector of rate constants $\mathbf{K}_{\mathrm{trans}}$ corresponding to the fluxes in (2.15) is easily represented as a Kronecker product, based on (2.14):

$$\mathbf{K}_{\mathrm{trans}} \equiv \Omega \otimes \mathbf{M} \quad \text{where}: \quad \Omega \equiv (\omega_1, \cdots, \omega_{N_{\mathrm{C}}})^T$$
$$\mathbf{M} \equiv (\mu_1, \cdots, \mu_{N_{\mathrm{S}}})^T \ . \tag{2.16}$$

We need to construct stoichiometry matrices that summarize the action of the $N_{\mathrm{C}} \cdot N_{\mathrm{S}}$ transfer processes in Ψ_{trans} on the $N_{\mathrm{L}} \cdot N_{\mathrm{S}}$ substances in $\mathbf{Y}$. First define two $N_{\mathrm{L}} \times N_{\mathrm{C}}$ matrices that summarize the source ($\mathbf{V}$) and destination ($\mathbf{W}$) locations of each link in the adjacency matrix:

$$\begin{aligned} \mathbf{V} &= \{v_{\mathrm{lc}}\} \quad , \quad v_{\mathrm{lc}} = \delta_{\,\mathrm{l},\sigma_c} \\ \mathbf{W} &= \{w_{\mathrm{lc}}\} \quad , \quad w_{\mathrm{lc}} = \delta_{\,\mathrm{l},\rho_c} \end{aligned} \quad (\forall) \ 1 \leq l \leq N_{\mathrm{L}} \ , \ 1 \leq c \leq N_{\mathrm{C}} \ . \tag{2.17}$$

The joint stoichiometry matrices corresponding to the transfer processes are Kronecker products of the above and $N_{\mathrm{S}} \times N_{\mathrm{S}}$ identity matrices,

$$\alpha_{\mathrm{trans}} \equiv \mathbf{V} \otimes \mathbb{I}_{\mathrm{N_S} \times \mathrm{N_S}} \quad ; \quad \beta_{\mathrm{trans}} \equiv \mathbf{W} \otimes \mathbb{I}_{\mathrm{N_S} \times \mathrm{N_S}} \ . \tag{2.18}$$

The net stoichiometry matrix for transport Γ_{trans} connects the transfer fluxes 2.15 to the corresponding rates of change in concentrations

$$\left(\frac{d\mathbf{Y}}{dt}\right)_{\text{trans}} = \Gamma_{\text{trans}} \cdot \Psi_{\text{trans}}(\mathbf{Y}) \tag{2.19}$$

$$\Gamma_{\text{trans}} \equiv (\mathbf{W} - \mathbf{V}) \otimes \mathbb{I}_{N_S \times N_S} \ . \tag{2.20}$$

Joint system: Finally, the joint rate vector $\Phi_{\text{tot}} = (\Phi_{\text{mol}}^T, \Psi_{\text{trans}}^T)^T$ obtained by concatenating the molecular and transfer parts, defines a mass-action CRN as follows:

$$\frac{d\mathbf{Y}}{dt} = \Gamma_{\text{total}} \cdot \begin{pmatrix} \Phi_{\text{mol}}(\mathbf{Y}) \\ \Psi_{\text{trans}}(\mathbf{Y}) \end{pmatrix}$$

$$\text{with } \Gamma_{\text{total}} \equiv \left(\mathbb{I}_{N_L \times N_L} \otimes \Gamma_{\text{loc}} \mid (\mathbf{W} - \mathbf{V}) \otimes \mathbb{I}_{N_S \times N_S}\right) \tag{2.21}$$

with the reactions and rates (similar to (2.2)) defined by the left- and right stoichiometry matrices

$$\alpha_{\text{total}} = \left(\mathbb{I}_{N_L \times N_L} \otimes \alpha_{\text{loc}} \mid \mathbf{V} \otimes \mathbb{I}_{N_S \times N_S}\right)$$

$$\beta_{\text{total}} = \left(\mathbb{I}_{N_L \times N_L} \otimes \beta_{\text{loc}} \mid \mathbf{W} \otimes \mathbb{I}_{N_S \times N_S}\right) \tag{2.22}$$

and the rate constant vector

$$\mathbf{K}_{\text{total}} = \begin{pmatrix} \mathbf{K}_{\text{mol}} \\ \mathbf{K}_{\text{trans}} \end{pmatrix} = \begin{pmatrix} \mathbb{1}_{N_L \times 1} \otimes \mathbf{K}_{\text{loc}} \\ \Omega \otimes \mathbf{M} \end{pmatrix} \ . \tag{2.23}$$

2.3 Application to VEGF Signal Initiation

2.3.1 Application to VEGF Signal Initiation: Complete Local Model

As an application of the formalism for combining a chemical reaction network model with a spatial network of domains, consider the CRN model of VEGF signal initiation [21]. The complete model describes two receptor types, which can form homo- and heterodimers, resulting in 16 chemical species and 24 reactions.

For the complete homodimer case (see Figure 2.1, let us consider the VEGF system containing only one type of receptor and ligand. The local

reactions possible are as follows:

$$\mathsf{R} + \mathsf{R} \underset{2}{\overset{1}{\rightleftharpoons}} \mathsf{R} \cdot \mathsf{R} \ ; \ \mathsf{V} \cdot \mathsf{R} + \mathsf{R} \underset{4}{\overset{3}{\rightleftharpoons}} \mathsf{V} \cdot \mathsf{R} \cdot \mathsf{R}$$

$$\mathsf{V} + \mathsf{R} \cdot \mathsf{R} \underset{6}{\overset{5}{\rightleftharpoons}} \mathsf{V} \cdot \mathsf{R} \cdot \mathsf{R} \ ; \ \mathsf{V} \cdot \mathsf{R} \cdot \mathsf{R} \underset{8}{\overset{7}{\rightleftharpoons}} \Delta$$

$$\mathsf{R} \cdot \mathsf{V} \cdot \mathsf{R} \underset{10}{\overset{9}{\rightleftharpoons}} \Delta \ ; \ \mathsf{V} \cdot \mathsf{R} + \mathsf{R} \underset{12}{\overset{11}{\rightleftharpoons}} \mathsf{R} \cdot \mathsf{V} \cdot \mathsf{R}$$

$$\mathsf{V} + \mathsf{R} \underset{14}{\overset{13}{\rightleftharpoons}} \mathsf{V} \cdot \mathsf{R} \tag{2.24}$$
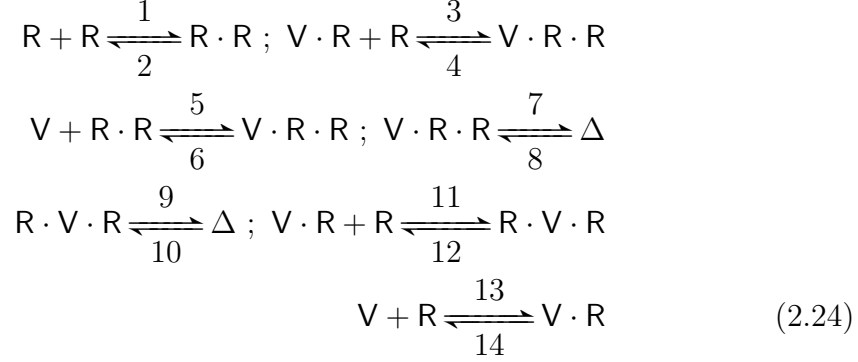
This yields seven local species and fourteen local reactions. The local species vector is

$$\mathbf{X}_{\mathrm{loc}} = (R, R \cdot R, V \cdot R, V \cdot R \cdot R, \Delta, R \cdot V \cdot R)^T \quad . \tag{2.25}$$

The species are as follows: there is one type of receptor which can be a free monomer R or may be ligand-bound $\mathsf{V}{\cdot}\mathsf{R}$, and receptors can bind to each other, and may form dimers without ligand $\mathsf{R}{\cdot}\mathsf{R}$. The VEGF ligand V has two receptor binding domains, thus we distinguish three different complexes that consist of two receptors and a ligand: $\mathsf{V}{\cdot}\mathsf{R}{\cdot}\mathsf{R}$ (ligand bound to one receptor, receptors bound to each other), $\mathsf{R}{\cdot}\mathsf{V}{\cdot}\mathsf{R}$ (receptor dimer with ligand bound to only on of them) and the fully bound complex Δ. It is generally believed that the $R - V - R$ structure is the key element in facilitating cross-activation of VEGF receptors, and the combined amount of $\mathsf{R}{\cdot}\mathsf{V}{\cdot}\mathsf{R}$ and Δ is identified with the "signal" output of the model.

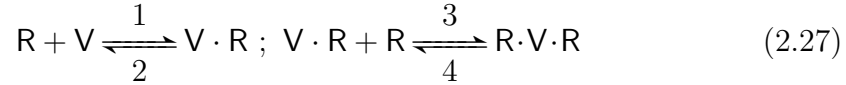The incoming and outgoing stoichiometric coefficient matrices α_{loc}, and

β_{loc} of the reaction network are as follows:

$$\alpha_{\mathrm{loc}} = \begin{bmatrix} 2 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

$$\beta_{\mathrm{loc}} = \begin{bmatrix} 0 & 2 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \qquad (2.26)$$

The steady state of the system defined above (and the complete model in [21]) has been investigated in [24] and a version of the (2.24) model with two domains was studied in [12]; however, we were unable to provide a definite answer regarding bistability.

2.3.2 Application to VEGF Signal Initiation: LID Only

Due to the number of species and reactions of the complete VEGF system (which has 7 species and 7 reversible reactions before including domain information), consider the following simplified Ligand Induced Dimerization (LID) only system (this simplification will allow for analytical calculations):

$$\mathsf{R} + \mathsf{V} \underset{2}{\overset{1}{\rightleftharpoons}} \mathsf{V} \cdot \mathsf{R} \; ; \; \mathsf{V} \cdot \mathsf{R} + \mathsf{R} \underset{4}{\overset{3}{\rightleftharpoons}} \mathsf{R} \cdot \mathsf{V} \cdot \mathsf{R} \qquad (2.27)$$

LID refers to ligand-induced dimerization, the mechanism where receptors can only dimerize in the presence of ligand following the sequence in (2.27), which remains after removing from (2.24) all reactions and species with direct $\mathsf{R} - \mathsf{R}$ binding. As a further simplifying hypothesis, we assume the VEGF ligand V is present at a fixed concentration V_0, so that the VEGF binding / dissociation reactions are represented as a reversible first order reaction,

with an effective rate constant derived from V_0 and the second order rate constant $k_{on,VR}$ of [21].

$$R \xrightleftharpoons[2']{1'} V \cdot R \ ; \ \ \varphi_1 = k_{eff}[R] \ , \ \ k_{eff} = V_0 k_{on,VR} \tag{2.28}$$
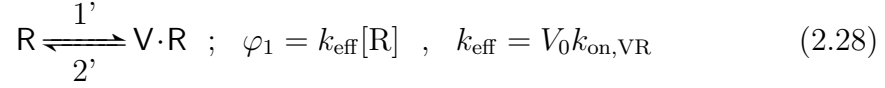
This system has species vector containing all three species seen in the reactions, the receptor, the receptor VEGF complex, and RVR, i.e. the signaling complex

$$\mathbf{X}_{loc} = (R, V \cdot R, RVR)^T \tag{2.29}$$

We read the local stoichiometric coefficients from the reactions. To create the local stoichiometric matrix, α_{loc}, we arrange a matrix with a number of rows representing the local species in order of appearance in the species vector above (2.29), and columns representing the reactions (2.28). For each column, we look at the inputs required for that individual reaction and place the stoichiometric coefficient in the appropriate row.

$$\alpha_{loc} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{2.30}$$

The same is done to construct the β_{loc} except instead of inputs we focus on the outputs of the reactions.

$$\beta_{loc} = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{2.31}$$

The final step in converting the local system into usable information is to construct the local rate vector by placing the rate constants into a vector in the same order as α and β were constructed.

$$\mathbf{K}_{loc} = (k_1, k_2, k_3, k_4)^T \tag{2.32}$$

To illustrate the transfer reactions, let us assume we have two identical locations that are interconnected. The adjacency matrix A is as follows:

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \; ; \; L = (1, 2) \tag{2.33}$$

Now we construct the total species vector, which is simply the local species vector repeated two times.

$$\mathbf{Y} = (R_1, V \cdot R_1, \Delta_1, R_2, V \cdot R_2, \Delta_2)^T \tag{2.34}$$

Using the Kronecker product we will expand the local alpha matrix, α_{loc} to α_{mol}.

$$\alpha_{\text{mol}} = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{2.35}$$

This is simply repeating α_{loc} along the block diagonal of the identity matrix of size two which is N_{L}. Repeating this process for β yields

$$\beta_{\text{mol}} = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \tag{2.36}$$

Finally we construct the total local rate vector

$$\mathbf{K}_{\text{mol}} = \mathbb{1}_{2 \times 1} \otimes \mathbf{K}_{\text{loc}} = (k_1, k_2, k_3, k_4, k_1, k_2, k_3, k_4)^T \tag{2.37}$$

We can see the practical way to construct the transfer pieces is to read the adjacency matrix, A from left to right and top to bottom. Reading it this way the first connection is in the top middle and is a path from the first

location to the second, the next is a connection from the second to the first. Using this we construct the transfer pieces of the example as follows:

$$\alpha_{\text{trans}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad ; \quad \beta_{\text{trans}} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \quad (2.38)$$
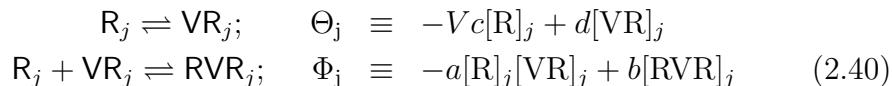
Assuming that the mobility rate vector for each species is given by $\mathbf{M} = (\mu_1, \mu_2, \mu_3)^T$, and that each location is identical, then the complete rate vector is simply $\mathbf{M}$ repeated twice.

$$\mathbf{K}_{\text{trans}} = (\mu_1, \mu_2, \mu_3, \mu_1, \mu_2, \mu_3)^T \quad (2.39)$$

The only remaining step is to put the molecular and transfer pieces together, and will be omitted. This is the LID only system which the analytics in the following section are based on as well as figures (2.5-2.6)

2.4 Analytical Equilibrium Calculations (Of the LID system)

Chemical transformations occur in all domains. There are two types of reactions (here to distinguish the analytical calculations from the previous section and for simplicity I renamed the rate constants to a, b, c, and d respectively):

$$\begin{aligned} \mathsf{R}_j &\rightleftharpoons \mathsf{VR}_j; & \Theta_{\mathsf{j}} &\equiv -Vc[\mathrm{R}]_j + d[\mathrm{VR}]_j \\ \mathsf{R}_j + \mathsf{VR}_j &\rightleftharpoons \mathsf{RVR}_j; & \Phi_{\mathsf{j}} &\equiv -a[\mathrm{R}]_j[\mathrm{VR}]_j + b[\mathrm{RVR}]_j \end{aligned} \quad (2.40)$$

Species can transfer between *certain* domain pairs. In this model, the only allowed transfers are between the "normal" domain (labeled by $j = 0$) and the other, "HD" domains. All three species may cross the boundary between an attractive domain and the rest of the membrane. The flux one direction is proportional to the concentration on the source side, with a universal constant, modulated by:

- A permeability factor (proportional to the length of the boundary)

- A mobility factor (dimers move at a reduced rate)

- The direction of the flux (transport into a high density domain is favored by a factor $\alpha > 1$)

All of these factors were included in $\mathbf{M} = (\mu_1, \mu_2, \mu_3)^T$. We also assume that $\mu_1 = \mu_2$ (i.e. both species of monomers have the same mobility) and that the mobility of dimers are proportional to monomers (i.e. $\mu_3 = \delta \cdot \mu_1$). To distinguish the analytical calculation from the previous section, I changed μ_j to v_j for this calculation.

2.4.1 Chemical Reaction Network Theory Applied to the LID System

Before continuing with analytical computations of the LID system, I attempted to apply some results from CRNT. With special thanks to Casian Pantea, I compared the complete VEGF system with one domain, the LID system, and the complete VEGF system with two domains to the Deficiency One Theorem proven by Feinberg in 1995 [18] (stated in the Appendix see 4.3). Using software by Dr. Pantea (available at https://reaction-networks.net/control/), it was determined that none of the CRNs satisfy the conditions for the Deficiency One Theorem, which states that if the deficiency of the CRN is 1 and the deficiency of the linkage classes add to 1 then there is a unique equilibrium if it exist, or (Theorem 6.1.1)[19] which gives uniqueness and existence conditions for deficiency zero CRNs if they are weakly reversible.

The terms used are as follows: $\mathcal{L}$ is the set of linkage classes, δ is the deficiency of the CRN, δ_i is the deficiency of linkage class i, m is the number of complexes, l is the number of linkage classes, and s is the span of stoichiometric matrix.

The details of the complete VEGF system with one domain:

$$\mathcal{L} = \{\{R_0, VR_0\}, \{R_0 + R_0, VRR_0, RVR_0, \Delta_0\}\}$$
$$\text{rank}(\Gamma) = 5$$
$$\delta = m - l - s = 8 - 2 - 5 = 1$$
$$\delta_1 = 2 - 1 - 1 = 0$$
$$\delta_2 = 4 - 1 - 3 = 0 \tag{2.41}$$

33

The details of the complete VEGF system with two domains:

$$\mathcal{L} = \{\{R_0, VR_0, R_1, VR_1\},$$
$$\{R_0 + R_0, VRR_0, RVR_0, \Delta_0, R_1 + R_1, VRR_1, RVR_1, \Delta_1\}\}$$
$$\text{rank}(\Gamma) = 11$$
$$\delta = m - l - s = 16 - 2 - 11 = 3 \qquad (2.42)$$

The details of the LID system with two domains:

$$\mathcal{L} = \{\{R_0, VR_0, R_1, VR_1\}, \{R_0 + VR_0, R_1 + VR_1, RVR_0, RVR_1\}\}$$
$$\text{rank}(\Gamma) = 5$$
$$\delta = m - l - s = 8 - 2 - 5 = 1$$
$$\delta_1 = 4 - 1 - 3 = 0$$
$$\delta_2 = 4 - 1 - 3 = 0 \quad (2.43)$$

As we can see from 2.41 2.42 and 2.43 neither of theorems apply, so I use analytics to calculate the equilibrium.

2.4.2   Calculations Continued

We define the transfer fluxes for any number of HD domains:

$$\begin{aligned}
\Psi_{\text{R,j}} &= v_j(-\alpha[\text{R}]_0 + [\text{R}]_j) \\
\Psi_{\text{VR,j}} &= v_j(-\alpha[\text{VR}]_0 + [\text{VR}]_j) \\
\Psi_{\text{RVR,j}} &= v_j\delta(-\alpha[\text{RVR}]_0 + [\text{RVR}]_j)
\end{aligned} \qquad (2.44)$$

Specifically when the number of high density domains is one, this becomes:

$$\begin{aligned}
\Psi_{\text{R,1}} &= v_1(-\alpha[\text{R}]_0 + [\text{R}]_1) \\
\Psi_{\text{VR,1}} &= v_1(-\alpha[\text{VR}]_0 + [\text{VR}]_1) \\
\Psi_{\text{RVR,1}} &= v_1\delta(-\alpha[\text{RVR}]_0 + [\text{RVR}]_1)
\end{aligned} \qquad (2.45)$$

For the transport equations, we must also take into account the change in area when a particle moves from one location to another as the concentration is inversely proportional to the area (call each A_j). The equations of motion

are therefore for any number of HD domains is:

$$\frac{d[\text{R}]_0}{dt} = \Theta_0 + \Phi_0 + \frac{1}{A_0} \sum_{j=1}^{n} \Psi_{R,j}$$

$$\frac{d[\text{VR}]_0}{dt} = -\Theta_0 + \Phi_0 + \frac{1}{A_0} \sum_{j=1}^{n} \Psi_{VR,j}$$

$$\frac{d[\text{RVR}]_0}{dt} = -\Phi_0 + \frac{1}{A_0} \sum_{j=1}^{n} \Psi_{RVR,j}$$

$$\frac{d[\text{R}]_j}{dt} = \Theta_j + \Phi_j - \frac{1}{A_j} \Psi_{R,j}$$

$$\frac{d[\text{VR}]_j}{dt} = -\Theta_j + \Phi_j - \frac{1}{A_j} \Psi_{VR,j}$$

$$\frac{d[\text{RVR}]_j}{dt} = -\Phi_j - \frac{1}{A_j} \Psi_{RVR,j} \tag{2.46}$$

For only one HD domain (which we can solve):

$$\frac{d[\text{R}]_0}{dt} = \Theta_0 + \Phi_0 + \frac{1}{A_0} \Psi_{R,1}$$

$$\frac{d[\text{VR}]_0}{dt} = -\Theta_0 + \Phi_0 + \frac{1}{A_0} \Psi_{VR,1}$$

$$\frac{d[\text{RVR}]_0}{dt} = -\Phi_0 + \frac{1}{A_0} \Psi_{RVR,1}$$

$$\frac{d[\text{R}]_1}{dt} = \Theta_1 + \Phi_1 - \frac{1}{A_1} \Psi_{R,1}$$

$$\frac{d[\text{VR}]_1}{dt} = -\Theta_1 + \Phi_1 - \frac{1}{A_1} \Psi_{VR,1}$$

$$\frac{d[\text{RVR}]_1}{dt} = -\Phi_1 - \frac{1}{A_1} \Psi_{RVR,1} \tag{2.47}$$

Some special items:

$$M_j = [\text{R}]_j + [\text{VR}]_j$$
$$E_j = [\text{R}]_j - [\text{VR}]_j$$
$$T_j = [\text{R}]_j + [\text{VR}]_j + 2[\text{RVR}]_j = M_j + 2[\text{RVR}]_j \tag{2.48}$$

35

This yields the conservation law

$$\sum_{j=1}^{n} T_j + T_0 = K \tag{2.49}$$

For the single HD domain case the conservation law becomes:

$$T_1 + T_0 = K \tag{2.50}$$

At equilibrium we also should note that:

$$\frac{-1}{A_j}(\Psi_{R,j} + \Psi_{VR,j} + 2\Psi_{RVR,j}) = 0$$
$$\alpha([R]_0 + [VR]_0 + 2\delta RVR_0) = [R]_j + [VR]_j + 2\delta[RVR]_j \tag{2.51}$$

At equilibrium, we add and subtract the equations of motion of $[R]_0$ and $[R]_j$ and $[VR]_0$ and $[VR]_j$ respectively, to arrive at

$$-A_0\Phi_0 = \sum_{j=1}^{n} \Phi_j$$
$$-A_0\Theta_0 = \sum_{j=1}^{n} \Theta_j \tag{2.52}$$

Manipulating the equations in (2.47), we arrive at:

$$-(c+d)\Theta_j + (d-c)\Phi_j + \frac{1}{A_j}(c\Psi_{R,j} - d\Psi_{VR,j}) = 0$$
$$-(c+d)\Theta_j + (d-c)\Phi_j + \frac{v_j}{A_j}(\alpha\Theta_0 - \Theta_j) = 0 \tag{2.53}$$

We will return to the general case later, but let us focus on the one HD domain case:

$$-(c+d)A_1\Theta_1 + (d-c)A_1\Phi_1 + v_1\alpha\Theta_0 - v_1\Theta_1 = 0$$
$$(c+d)A_0\Theta_0 + (c-d)A_0\Phi_0 + v_1\alpha\Theta_0 + \frac{v_1 A_0}{A_1}\Theta_0 = 0$$
$$(c+d+\frac{v_1}{A_1}+\frac{v_1\alpha}{A_0})\Theta_0 = (d-c)\Phi_0 \tag{2.54}$$

36

Let $\Omega = \frac{c + d + \frac{v_1}{A_1} + \frac{v_1 \alpha}{A_0}}{d - c}$ we arrive at:

$$\Omega \Theta_0 = \Phi_0 \tag{2.55}$$

Now we can attempt to solve the one HD case and will begin with removing Φ_0 from the equations:

$$
\begin{aligned}
\frac{d[\mathrm{R}]_0}{dt} &= (1 + \Omega)\Theta_0 + \frac{1}{A_0}\Psi_{R,1} \\
\frac{d[\mathrm{VR}]_0}{dt} &= (-1 + \Omega)\Theta_0 + \frac{1}{A_0}\Psi_{VR,1} \\
\frac{d[\mathrm{RVR}]_0}{dt} &= -\Omega\Theta_0 + \frac{1}{A_0}\Psi_{RVR,1}
\end{aligned}
\tag{2.56}
$$

Combining (31) and (33) with the conservation law and eliminating $[\mathrm{R}]_1$ and $[\mathrm{VR}]_1$:

$$\frac{A_1\alpha}{A_0}([\mathrm{R}]_0 + [\mathrm{VR}]_0 + 2\delta[\mathrm{RVR}]_0) + \frac{2A_1(1-\delta)}{A_0}[\mathrm{RVR}]_1 + T_0 = K \tag{2.57}$$

Now we can eliminate $[\mathrm{RVR}]_1$

$$[\mathrm{RVR}]_1 = \alpha[\mathrm{RVR}]_0 + \frac{A_0\Omega}{\delta v_1}\Theta_0 \tag{2.58}$$

Substituting again:

$$\frac{A_1\alpha}{A_0}([\mathrm{R}]_0 + [\mathrm{VR}]_0 + 2\delta[\mathrm{RVR}]_0) + \frac{2A_1(1-\delta)}{A_0}(\alpha[\mathrm{RVR}]_0 + \frac{A_0\Omega}{\delta v_1}\Theta_0) + T_0 = K$$

$$\frac{A_1\alpha}{A_0}([\mathrm{R}]_0 + [\mathrm{VR}]_0) + \frac{2A_1\alpha}{A_0}[\mathrm{RVR}]_0 + \frac{2A_1\Omega(1-\delta)}{\delta v_1}\Theta_0 + T_0 = K$$

Let:

$$
\begin{aligned}
K_{R,1} &= 1 + \frac{A_1\alpha}{A_0} \\
K_{\Theta,1} &= \frac{2A_1\Omega}{\delta v_1}
\end{aligned}
\tag{2.59}
$$

we arrive at

$$K_{R,1}[\mathrm{R}]_0 + K_{R,1}[\mathrm{VR}]_0 + 2K_{R,1}[\mathrm{RVR}]_0 + K_{\Theta,1}\Theta_0 = K \tag{2.60}$$

37

Now to eliminate $[\text{RVR}]_0$:

$$[\text{RVR}]_0 = \frac{a}{b}[\text{R}]_0[\text{VR}]_0 + \frac{\Omega}{b}\Theta_0$$

$$K = K_{R,1}[\text{R}]_0 + K_{R,1}[\text{VR}]_0 + 2K_{R,1}\frac{a}{b}[\text{R}]_0[\text{VR}]_0 + (K_{\Theta,1} + 2K_{R,1}\frac{\Omega}{b})\Theta_0$$

Let:

$$K_{\Theta,2} = K_{\Theta,1} + 2K_{R,1}\frac{\Omega}{b} \tag{2.61}$$

$$K = (K_{R,1} - cK_{\Theta,2})[\text{R}]_0 + (K_{R,1} + dK_{\Theta,2})[\text{VR}]_0 + 2K_{R,1}\frac{a}{b}[\text{R}]_0[\text{VR}]_0$$

For simplicity, one more substitution of constants:

$$
\begin{aligned}
K_{R,2} &= K_{R,1} - cK_{\Theta,2} \\
K_{VR,2} &= K_{R,1} + dK_{\Theta,2} \\
K_{R \cdot VR,2} &= 2K_{R,1}\frac{a}{b}
\end{aligned}
\tag{2.62}
$$

Leading to the following equation for $[\text{VR}]_0$ in terms of $[\text{R}]_0$:

$$[\text{VR}]_0 = \frac{K - K_{R,2}[\text{R}]_0}{K_{VR,2} + K_{R \cdot VR,2}[\text{R}]_0} \tag{2.63}$$

For the second equation:

$$
\begin{aligned}
[\text{R}]_1 &= \alpha[\text{R}]_0 - \frac{A_0}{v_1}(1+\Omega)\Theta_0 \\
[\text{VR}]_1 &= \alpha[\text{VR}]_0 + \frac{A_0}{v_1}(1-\Omega)\Theta_0
\end{aligned}
\tag{2.64}
$$

$$[\text{R}]_1 \cdot [\text{VR}]_1 = \alpha^2[\text{R}]_0 \cdot [\text{VR}]_0 \ + \ \frac{\alpha A_0}{v_1}(1-\Omega)[\text{R}]_0\Theta_0 - \frac{\alpha A_0}{v_1}(1+\Omega)\Theta_0 \cdot [\text{VR}]_0$$

$$- \frac{A_0^2}{v_1^2}(1+\Omega)(1-\Omega)\Theta_0^2 \tag{2.65}$$

We can again simplify the equation by renaming the constants:

$$Z_{R_0^2} = \frac{\alpha A_0}{v_1}(1 - \Omega)c + \frac{A_0^2}{v_1^2}(\Omega^2 - 1)c^2$$

$$Z_{VR_0^2} = \frac{-\alpha A_0}{v_1}(1 + \Omega)d + \frac{A_0^2}{v_1^2}(\Omega^2 - 1)d^2$$

$$Z_{R \cdot VR} = \alpha^2 + \frac{\alpha A_0}{v_1}(1 + \Omega)c - \frac{\alpha A_0}{v_1}(1 - \Omega)d - 2cd\frac{A_0^2}{v_1^2}(\Omega^2 - 1)$$

$$[\text{R}]_1 \cdot [\text{VR}]_1 = Z_{R_0^2}[\text{R}]_0^2 + Z_{VR_0^2}[\text{VR}]_0^2 + Z_{R \cdot VR}[\text{R}]_0 \cdot [\text{VR}]_0 \qquad (2.66)$$

Using $\Phi_1 = -\frac{1}{A_1}\Psi_{RVR,1}$ along with the conservation law, we arrive at:

$$-a[\text{R}]_1 \cdot [\text{VR}]_1 + b[\text{RVR}]_1 = \frac{-v_1\delta}{A_1}(-\alpha[\text{RVR}]_0 + [\text{RVR}]_1)$$

$$\frac{A_1 a}{v_1\delta}[\text{R}]_1 \cdot [\text{VR}]_1 + \alpha[\text{RVR}]_0 = (1 + \frac{A_1 b}{v_1\delta})[\text{RVR}]_1$$

$$X_{R_1 \cdot VR_1} = \frac{A_1 a}{v_1\delta}(1 + \frac{A_1 b}{v_1\delta})^{-1}$$

$$X_{RVR_0} = \alpha(1 + \frac{A_1 b}{v_1\delta})^{-1}$$

$$[\text{RVR}]_1 = X_{R_1 \cdot VR_1}[\text{R}]_1 \cdot [\text{VR}]_1 + X_{RVR_0}[\text{RVR}]_0 \qquad (2.67)$$

Recall:

$$[\text{RVR}]_1 = \alpha[\text{RVR}]_0 + \frac{A_0\Omega}{v_1\delta}\Theta_0 \qquad (2.68)$$

By substitution:

$$\alpha[\text{RVR}]_0 + \frac{A_0\Omega}{v_1\delta}\Theta_0 = X_{R_1 \cdot VR_1}[\text{R}]_1 \cdot [\text{VR}]_1 + X_{RVR_0}[\text{RVR}]_0$$

$$(\alpha - X_{RVR_0})[\text{RVR}]_0 = X_{R_1 \cdot VR_1}[\text{R}]_1 \cdot [\text{VR}]_1 - \frac{A_0\Omega}{v_1\delta}\Theta_0$$

$$Y_{R_1 \cdot VR_1} = X_{R_1 \cdot VR_1}(\alpha - X_{RVR_0})^{-1}$$

$$Y_{\Theta_0} = -\frac{A_0\Omega}{v_1\delta}(\alpha - X_{RVR_0})^{-1}$$

$$[\text{RVR}]_0 = Y_{R_1 \cdot VR_1}[\text{R}]_1 \cdot [\text{VR}]_1 + Y_{\Theta_0}\Theta_0 \qquad (2.69)$$

39

By substituting into the conservation law:

$$\begin{aligned}
K &= K_{R,1}[\text{R}]_0 + K_{R,1}[\text{VR}]_0 + 2K_{R,1}(Y_{R_1 \cdot VR_1}[\text{R}]_1 \cdot [\text{VR}]_1 + Y_{\Theta_0}\Theta_0) + K_{\Theta,1}\Theta_0 \\
Z'_{R_0} &= K_{R,1} - c(2K_{R,1}Y_{\Theta_0} + K_{\Theta,1}) \\
Z'_{VR_0} &= K_{R,1} + d(2K_{R,1}Y_{\Theta_0} + K_{\Theta,1}) \\
Z'_{R_0^2} &= 2K_{R,1}Y_{R_1 \cdot VR_1}Z_{R_0^2} \\
Z'_{VR_0^2} &= 2K_{R,1}Y_{R_1 \cdot VR_1}Z_{VR_0^2} \\
Z'_{R_0 \cdot VR_0} &= 2K_{R,1}Y_{R_1 \cdot VR_1}Z_{R_0 \cdot VR_0}
\end{aligned}$$

Finally arriving at our second equation:

$$Z'_{R_0}[\text{R}]_0 + Z'_{VR_0}[\text{VR}]_0 + Z'_{R_0^2}[\text{R}]_0^2 + Z'_{VR_0^2}[\text{VR}]_0^2 + Z'_{R_0 \cdot VR_0}[\text{R}]_0 \cdot [\text{VR}]_0 = K$$

(2.70)

Instead of solving the resulting 4th order equation analytically, it is solved numerically (graphically) in section 2.5.1.

2.5 Simulations of VEGF

The values of the local rate constants for the VEGF system were taken from a 2007 MacGabhann paper [21], however some of these numbers had to be converted from use for ODE/PDE and concentrational values to Gillespie rate constants. In the MacGabhann paper, the rates for dimerization were given with the units $cm^2 \cdot mol^{-1} \cdot s^{-1}$. The others had units of $M^{-1} \cdot s^{-1}$ and s^{-1}. To convert the units into Gillespie rates, for the first, the area and Avogadro's number had to be divided out, the next the volume and Avogadro's number, but the last required no change. This conversion was necessary as we are keeping track of the number of individuals instead of the concentration of those species.

2.5.1 Analytics and Validation

We calculated the equilibrium values for the system without ligand and immobile dimers. More over this obviously has only one positive solution, meaning there is a unique (positive) equilibrium. (See Figure (2.5)).

In order to validate the simulations, we computed the values of $[\text{R}]_0$ and $[\text{VR}]_0$ at equilibrium and compared this result to the simulation. The results
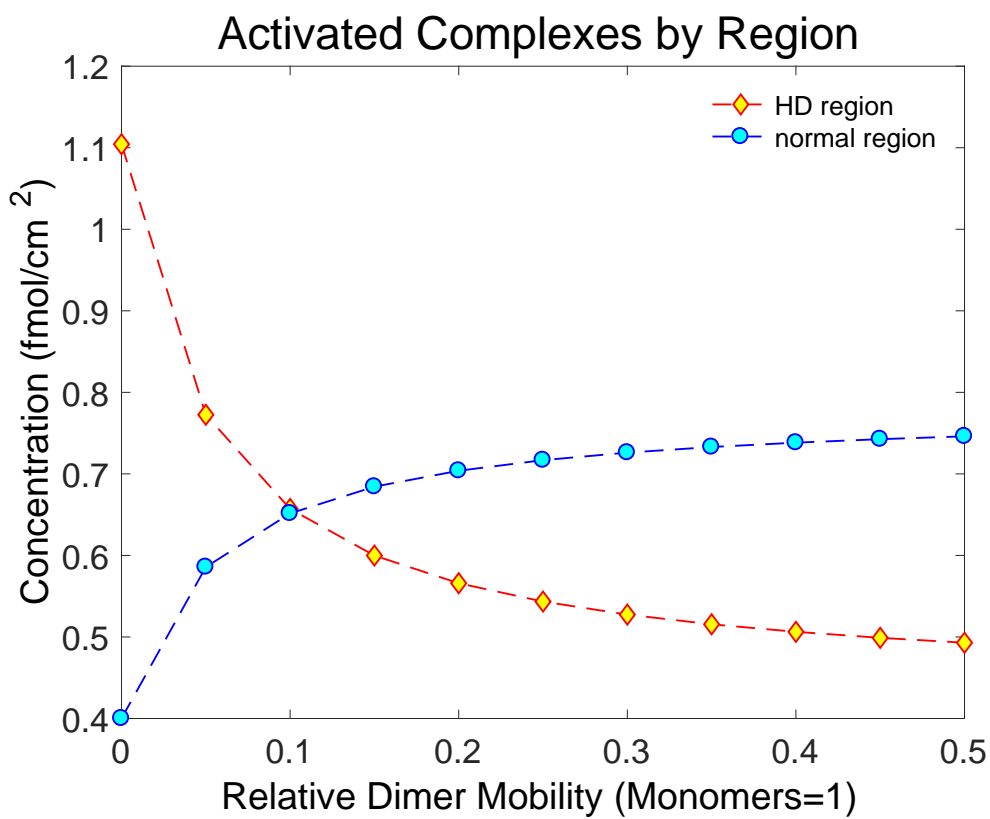
Figure 2.2: Effect of dimer mobility on signaling in the HD and normal sector. The figure uses the rate coefficients from the MacGabhann paper. This graph shows the effect of dimer mobility (one variable that is not present in the MacGabhann model) on the concentration of signaling complexes. The mobility is varied from completely immobile represented by 0 and half as mobile as a monomer (a reasonable place to stop). We see the concentration of signaling complexes in yellow increases in the attractive or High Density Domain as the dimer mobility decreases.

Figure 2.3: Effect of ligand concentration on signaling in the HD and normal sector. The figure again uses the rate coefficients from the MacGabhann paper. This graph on log scale shows that increasing the concentration of ligand increases the signaling complexes in both the attractive and normal domain. The ligand concentration does not reach a high enough level to achieve high dose inhibition.
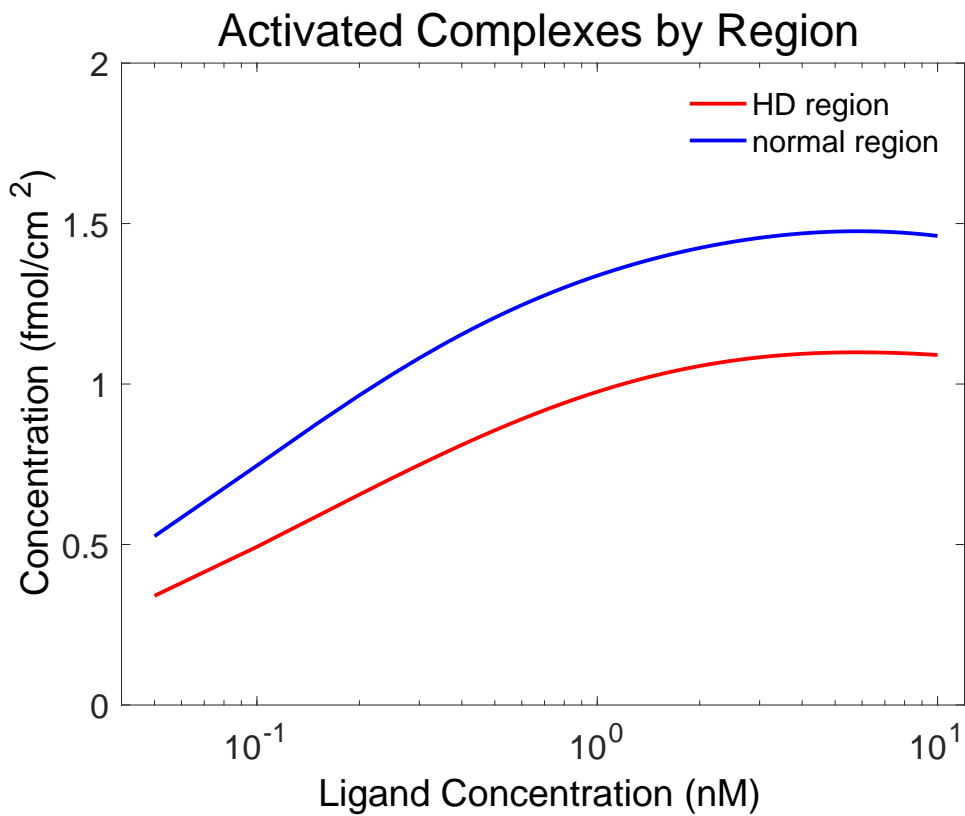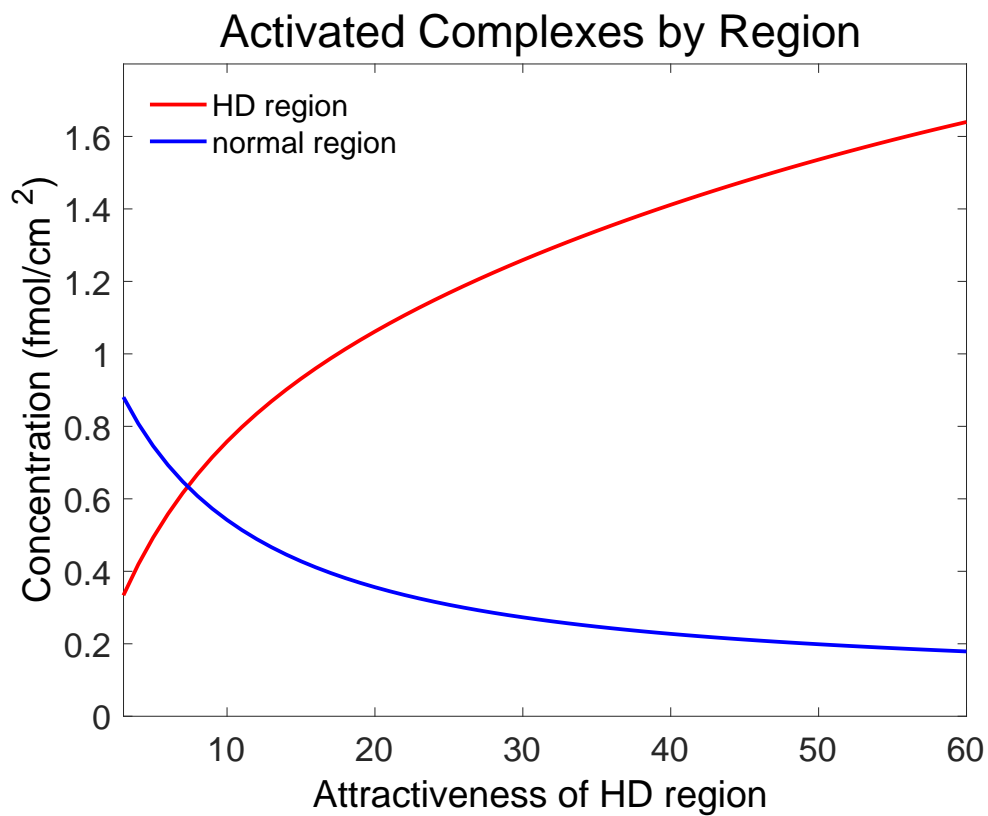
Figure 2.4: Effect of domain attractiveness on signaling in the HD and normal sectors. The figure again uses the rate coefficients from the MacGabhann paper. This graph shows the effect of increasing attractiveness of the attractive region on signaling complexes.

can be seen in Figure (2.6). The points are values of the simulation as time increases towards infinity. The red and blue graphs are the results of the analytical calculations. We do arrive at two equations with two unknowns; however, instead of solve that system we will graph it (using what we set for default values) for comparison to on ODE simulation.

$$[VR]_0 = \frac{K - K_{R,2}[R]_0}{K_{VR,2} + K_{R \cdot VR,2}[R]_0}$$

$$Z'_{R_0}[R]_0 + Z'_{VR_0}[VR]_0 + Z'_{R_0^2}[R]_0^2 + Z'_{VR_0^2}[VR]_0^2 + Z'_{R_0 \cdot VR_0}[R]_0 \cdot [VR]_0 = K$$

2.6   Conclusions

The first part of thesis is the construction and analysis of a model that combines a given Chemical Reaction Network with a spatial network of domains. The starting point is a model of VEGFR signal initiation due to MacGabhann and Popel [21] (and previous work cited therein). The choice of this biological system is due to its medical relevance. VEGF is responsible for the development of blood vessels and has been proposed as a target for tumor suppression strategies.

Experimental data (static TEM images revealing clustering of receptors, highly unlikely for random placement and single particle tracking that points to regions that 'trap' receptors by slowing them down) indicates that the cellular membrane is not homogenous. Rather than a uniform 'two dimensional liquid' that allows free diffusion, it exhibits a network of domains (presumably induced by lipid rafts, cytosketal elements and others) that tend to attract and group VEGF receptors. Thus, I was prompted to modify the Popel and MacGabhann model [21] to include this spatial information. I included the spatial structure by developing the microdomain/boxes model. In the resulting model each spatial domain is represented by a well-mixed "compartment". Each compartment (microdomain/box) contains a complete copy of every species and reaction with transfer "reactions" to represent diffusion between the compartments. The resulting model is still a chemical reaction network that can be analyzed in the context of CRNT and simulated both as a system of ordinary differential equations or following a stochastic simulation algorithm such as the Gillespie SSA. This is an economical way to incorporate the presence of nontrivial spatial structures, which is computationally less intense than going to a fully spatial model. Because of this
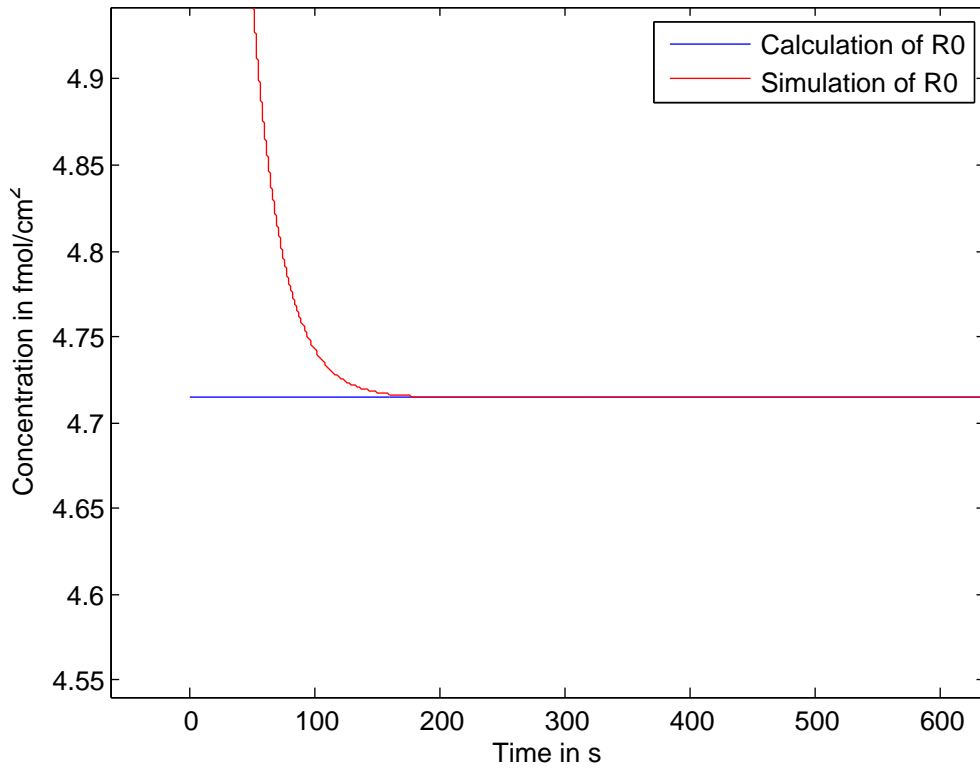
Figure 2.5: Convergence to equilibrium. A check that our analytical computations match with the simulation at equilibrium the concentration of R_0, the receptor in the normal domain. The analytical equilibrium value is the line shown in blue and the results of an ODE simulation are shown in red.
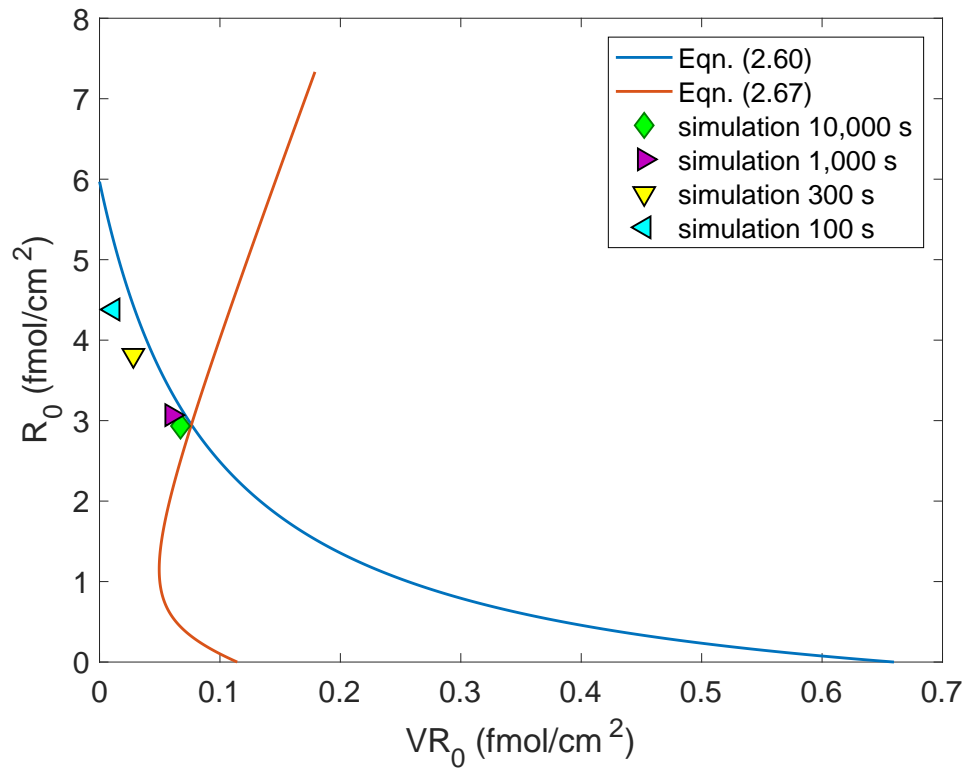
Figure 2.6: Convergence to equilibrium, parameter space view. Another check that our analytical computations match with the simulation. The red and blue curves are from the analytical calculations with the equilibrium value at the intersection of the two graphs and the points are several time points of an ODE simulation with convergence toward the equilibrium.

simpler model, analytical computations and insight are possible, such as the equilibrium calculation performed in the text.

The main technical contribution is the framework we use to derive the joint CRN model, as a composition of the local CRN model and the spatial network. From the knowledge that the VEGFR is monovalent while the receptor is bivalent, combined with Mass Action Rate Laws we can develop the local stoichiometric matrices as done by Popel and MacGabhann (they also developed the rate constants we used for the reactions) [21].

Using the Kronecker Product, the local stoichiometric matrices, and location information, I built the complete stoichiometric matrices for the coarse grained model. In order to check my simulations, equilibrium calculations were made and compared to the results from simulation at the final time steps (as time approached "infinity").

After validating the simulation, we can allow values to vary and observe the effects on signaling strength. We can see from Figures(2.2 2.3 2.4) that:

1) Lower dimer mobility favors signaling in the attractive region as with lower dimer mobility, more monomers would become trapped in the attractive region, become dimers and then be less likely to exit

2) Increased ligand (we did not reach high dose inhibition) increases the signal in all domains, but more so in the attractive region and

3) Increasing the attractiveness beyond 8 fold leaves a higher actual concentration of signaling complex in the attractive region.

We should note that as time increasing we would expect a lower dimer mobility as more proteins bind to the activated interior increasing the size of the signaling complex as well as slowing it. This would also suggest that a dimer breaking apart into its monomer components becomes exceedingly less likely as time progresses. This model also ignores internalization which would turn a large signaling complex back to only monomers.

Chapter 3

Deriving effective non-spatial kinetics from a spatial model of ErbB2 / ErbB3 dimerization

3.0.1 Biology of EGF

EGFR is also a member of the receptor-tyrosine kinase family, one of the largest families of membrane receptors [58] and one of the first growth factors discovered [4]. The epidermal growth factor receptor family consists of the following receptors: EGFR (ErbB1), ErbB2, ErbB3 and ErbB4 [60]. EGF receptors, EGFR, tend to be transmembrane proteins [38]. ErbB3 is dependent on heterodimerizing with an ErbB receptor for phosphorylation and increasing its weak kinase activity. The preferred dimerizing partner for ErbB3 is ErbB2. Known ligands for ErbB3 include Neuregulin -1 and Neuregulin - 2. [62] Because of the nature of heterodimers, many different signaling outcomes are possible [4]. "EGFR plays important roles in the regulation of cell growth and development." [58] The ERBB family also has roles in "normal embryonic development, angiogenesis, metastasis, cell proliferation and apoptosis resistance." [29] Activated EGFR also prevents apoptosis (cell death) [58] and play a role in cellular differentiation [4]. Proper EGF signaling helps control proper haploid cell generation [11].

EGF acts as mitogens and motogens which transform epithelial cells into carcinomas[38]. EGF also plays a role in lung cancer [30], and in cell proliferation and migration [34]. Getfitinib is given to inhibit EGF in lung cancer and breast cancer [10]. Getfitinib is also used in combination with methotrexate

to treat ectopic pregnancy [10]. EGFR is "associated with poor prognosis of breast cancer patients." [38] It is known that EGF is correlated to "malignant transformation and all the steps of the breast cancer metastatic cascade."[57] This is partial because of a correlation between EGF and endocrine resistant breast cancer.[57] Over expression of ERBB2 has been linked to gastric cancer [29]. Not only is the EGF family involved in tumor progression, but also survival under the harsh conditions of chemotherapy and radiotherapy [38]. EGFR prevents the maturation of tumor suppressing microRNAs (miRNA) [38].

The EGFR family of receptors is expressed in many types of cell [4]. Similarly to VEGF the receptors must not only find the target ligand, but also find another receptor of the appropriate type in order to signal. Unlike VEGF which has a bivalent ligand, EGF has only monovalent ligands. This means that there is no high dose inhibition of EGF unlike VEGF (high dose inhibition in VEGF is caused by too many receptors already bound to ligands to signal).

Like VEGFR, EGFR has binding proteins that have a Src Homology 2 (SH2) binding domain [40]. Several downstream pathways are the target of Ras [40]. "Data suggest that the physiological outcome of tyrosine kinase signaling strongly depends on the timing, duration, and amplitude of activation of signaling components." [40] This gives even more importance for finding a suitable model for EGFR-EGF binding. Subtle differences in signals could possibly create different responses [40]. miRNA regulation may be influenced by growth factors[38]. For example, changes in expression of miRNAs were seen in healthy cells after exposure to EGF [38]. This is complicated by the fact that miRNAs also influence the expression of growth factor receptors [38]. In this paper I will not include these complications to the model, but include mentioning them because of the medical relevance. It also suggests that other factors besides the SH2 domain must be responsible for cell to tell a VEGF signal from an EGF signal. I have not seen any facts in the literature searches explain how the cell can tell the two signals apart.

3.1 Generating the well-mixed parameters for the ErbB2 and ErbB3 system

3.1.1 Spatial stochastic model: Brownian motion simulation and binding radius

In the discussion below, we refer to a spatial simulation algorithm, as implemented in [39] and references therein. The motion simulator and triggering of reactions is essentially identical to Smoldyn [3]. Smoldyn is widely used in the literature, although there is an array of other methods [49] [17] and available software [13] [41] . We will focus on particle movement simulated through a first principles Brownian motion approach, dimerization and dissociation; we will not discuss zero- and first order reactions and other features implemented in the Smoldyn software. The Smoldyn algorithm relies on a fixed time step Δt; particles are represented by geometric points that move at each update according to Brownian motion and participate in reactions according to rules as described below.

We will start our discussion of the Smoldyn simulation with dimerization and dissociation and finish with Brownian motion. In the simulation, dimerization is caused when two particles diffuse via Brownian motion (more on this later) within a binding radius (BR) of each other. The BR (denoted ρ_b) is determined using the kinetics of the system for each reaction type and is such that the behavior of the simulation matches experimental data [3]. This means that whenever two eligible particles are within the BR of each other dimerization will occur. In order to save computational time, the simulation ignores the probability of the reaction occurring provided the eligible particles meet (the BR already took this probability into account).

Dissociation is controlled by a decay rate k_{off}. As every first order reaction, dissociation is triggered stochastically at each full update; the probability is $\Delta p = 1 - \exp\left(-k_{\mathrm{off}}\Delta t\right) \approx k_{\mathrm{off}}\Delta t$ for each existing dimer. The only other consideration for dissociation is the unbinding radius (UBR, denoted ρ_{ub}). The UBR, denoted the distance dissociated particle are placed apart, is usually around 5 times greater than the BR and is used to prevent the particles from immediately rebinding [3].

The final item the simulation uses is a first-principles Brownian motion to simulate the movement (diffusion) of particles. Brownian motion is the core element in the Smoldyn simulation. It allows all of the parti-

cles to move around the simulation space by moving each molecule following $[x(t + \Delta t), y(t + \Delta t)] = [x(t), y(t)] + [\Delta x, \Delta y]$. The displacements are derived from a normal distribution with variance $\sigma^2 = 2D_0\Delta t$. The collision rate is influenced by the choice of the time step. If the typical displacement $\sqrt{2D_0\Delta t}$ exceeds BR, some collisions will be missed, reducing the reaction rate.

In this paper we are attempting to create a well-mixed model of the dimerization of ErbB2 and ErbB3. We will use Smoldyn simulation, characterized by $D_0, k_{\text{off}}^{(\text{sim})}, \Delta t$, BR and UBR, to compare to the solution of the well-mixed model with yet undetermined rate constants $k_{\text{on}}^{(\text{eff})}$ and $k_{\text{off}}^{(\text{eff})}$.

3.1.2 ODE model: well-mixed, continuous and deterministic

The traditional (or vastly more widespread) model for bio-molecular processes relies on well-mixed chemical reactions with continuous dynamics. In this approach, a generic dimerization-dissociation reaction is represented by the chemical equation $A + B \rightleftharpoons C$. The dynamics is described in terms of the concentration of each substance [A],[B],[C]; their time evolution is determined by rate laws that set the flux through each reaction.

$$\frac{d[A]}{dt} = \frac{d[B]}{dt} = -\frac{d[C]}{dt} = -\varphi_{\text{A+B}\rightarrow\text{C}}\left([A], [B], [C]\right) + \varphi_{\text{C}\rightarrow\text{A+B}}\left([A], [B], [C]\right)$$

(3.1)

Rate laws used to describe signaling networks represent intricate reaction mechanisms, but are ultimately are derived from mass action rate laws, which state that the rate of a simple reaction is proportional to the product of the amounts of the reacting substances. The proportionality constants $k_{\text{on}}, k_{\text{off}}$ are commonly used to characterize the rate of a given reaction:

$$\varphi_{\text{A+B}\rightarrow\text{C}} = k_{\text{on}}[A][B] \; ; \; \varphi_{\text{C}\rightarrow\text{A+B}} = k_{\text{off}}[C] \quad \Rightarrow$$
$$\frac{d[A]}{dt} = \frac{d[B]}{dt} = -\frac{d[C]}{dt} = -k_{\text{on}}[A][B] + k_{\text{off}}[C] \quad (3.2)$$

The ODE model relies on two implicit simplifications. First, the state of the system is sufficiently described by the amounts of each substance; spatial

51

inhomogeneities are either negligible or can be abstracted away[1]. We will refer to this as the *well mixed assumption*. Second, the amount of each substance may be represented by a (non-negative), continuous quantity; this becomes untenable when the number of molecules involved becomes small. In a cell diameter of $d \approx 10\,\mu\text{m}$ and a concentration of $[A] = 1\,\text{nM}$ there are only a few hundreds:

$$N \approx \frac{\pi}{6} \times 10^3 \frac{\mu\text{m}^3}{\text{cell}} \times 10^{-9}\,\text{M} \times 6 {\cdot} 10^{23} \frac{\text{molec}}{\text{litre}}/\text{M} \times 10^{-15} \frac{\text{litre}}{\mu\text{m}^3} \approx 314 \frac{\text{molec}}{\text{cell}}$$

(3.3)

ODE models of signaling networks give rise to complex dynamical systems requiring specialized tools for simulation (rule based modeling [61] is often necessary to deal with the combinatorial explosion of the number of species) and theoretical analysis [5] [32]. On the experimental side, individual processes are typically characterized by kinetic parameters related to the traditional rate constants.

Thus, while spatial and stochastic models are a more faithful representation of the physical reality, their generalized use in signaling models would raise the cost of simulations and further complicate the task of extracting biologically relevant insight. Given the uncertainty on kinetic parameters and the nonlinear dynamics involved, the added value might be limited. Instead, spatial simulations can focus on specific aspects where detailed experimental insight exists. We need consistent abstractions that summarize this detailed picture in terms of effective kinetic parameters that relate to larger scale models and the vast majority of accumulated experimental information.

The goal of this paper is to investigate the relation between a popular type of spatial simulation (based on Smoldyn [3][2] and implemented by this group [39] [45]), ODE models, and non-spatial stochastic simulations. We will not discuss numerical simulation of ODE systems, which is a mature discipline with well-established methods [23]. We will make use of the exact analytical solution of dimerization-dissociation with mass action rates [26].

---

[1]Abstractions may remain valid if significant inhomogeneities exist but their effect on the total amount of substance (or average concentration) can be captured by appropriately [re]defined rate laws.

3.1.3 Non-spatial stochastic model: Gillespie simulations and correspondence principle

Non-spatial stochastic simulations of mass-action reaction systems are fairly standard. We follow the approach of the Gillespie SSA [22]. The amount of each substance is represented by the respective number of molecules, and individual reaction events are stochastically occurring transitions in the resulting discrete state space. Reactions of a given type are triggered by Poisson processes whose rate (probability per time or propensity γ_{reaction}) is equivalent to the mass-action rate law.

The correspondence principle is that the rate *per particle* for each of the reacting species in a reaction equals the effective time constant for that species resulting from the continuous rate law. For dissociation (and any other first order reaction), the stochastic rate constant equals the macroscopic one

$$\gamma_{C \to A+B} = k_{\text{off}} \cdot N_C \leftrightarrow \frac{\gamma_{C \to A+B}}{N_C} = \frac{\varphi_{C \to A+B}}{[C]} = \frac{k_{\text{off}} \cdot [C]}{[C]} = k_{\text{off}} \qquad (3.4)$$

For second- and higher order reactions, a volume factor must be included. For dimerization, assuming $[A] = N_A V_0$ and $[B] = N_B V_0$ we have[2]

$$\frac{\gamma_{A+B \to C}}{N_A} = \frac{\varphi_{A+B \to C}}{[A]} = \frac{k_{\text{on}} \cdot [A] \cdot [B]}{[A]} = \cdots$$

$$k_{\text{on}} \cdot [B] = k_{\text{on}} \cdot \frac{N_B}{V_0} \to \gamma_{A+B \to C} = \frac{k_{\text{on}}}{V_0} \cdot N_A \cdot N_B \qquad (3.5)$$

3.1.4 Biological model system

This work was motivated by our investigation [39] [50] of the EGF ligand / receptor family, in particular ErbB2 and ErbB3, which are important for their role in cancer [15]. The membrane bound receptors are capable of both homodimerization and heterodimerization [50]. The rationale of the developing the spatial model is to connect to single particle tracking (SPT) [53] data. In SPT, labeled molecules are directly observed diffusing across the membrane and dimerizing. The spatial model was developed using this

---

[2]Note that when the reacting substances occupy different spatial realms (such as membrane bound receptors binding ligand from the extracellular space), the reaction rates for the two reacting species are in different units, i.e. mol per area versus mol per volume.

method in the context of a recent study of the effect of membrane domains on signaling [39]. ErbB2 and ErbB3 are receptor-tyrosine kinases (RTK) [50]; that means dimerization is required for phosphorylation [58]. Moreover, in the ErbB2/ErbB3 system, heterodimerization is required for phosphorylation/signal initiation [50].

3.2 Methods

3.2.1 Spatial stochastic simulation

In this work, we use a similar algorithm to the one in Smoldyn, as suggested by Andrews and Bray [2] [45]. This is a fixed time step algorithm with a Brownian Motion simulator for particle movement. Molecular species are represented as geometric points. Binary reactions (dimerization) are triggered by collisions between participating particles; all other transformations (dissociation in the present case) are triggered stochastically, consistent with the propensity (rate) associated with the espective process. Each binary reaction is characterized by an associated "binding radius" (ρ_B); when two potential partners are within this distance, a reaction happens with certainty. To avoid instant recombination, particles resulting from dissociation are placed at a distance called the "unbinding radius" $\rho_U > \rho_B$. The actual simulation proceeds with one-particle updates when one particle (receptor) is selected randomly; it is first given a Brownian displacement corresponding to Δt; then we check for collisions and implement the resulting dimerization if it is the case; if no binary reaction occurs, an intrinsic reaction is implemented with probaiblity $\gamma \Delta t$. The nominal time step in the simulation Δt corresponds to a full update of the state of all particles, therefore the time counter is advanced by $\Delta t/N_{\text{particles}}$ for each one-particle update. The parameters are listed in the table 3.1.

3.2.2 Well mixed stochastic simulation

We also make use of the Gillespie Algorithm. We in particular use the Next Event Method, where the total probability is used to generate the time any event will occur then randomly select the event that occurs, because it is less computationally intense compared to the original Gillespie. For the general case we had to generate the stoichiometric matrices using a rules

Table 3.1: Smoldyn-like algorithm parameters

Diffusion coefficient (monomers)	$D_{\mathrm{mon}} = 3.1 \cdot 10^{-2}\frac{m^2}{s}$
Diffusion coefficient (dimers)	$D_{\mathrm{dim}} = 0.5 \cdot D_{\mathrm{mon}} = 1.55 \cdot 10^{-2}\frac{m^2}{s}$
Dissociation rate constant	$k_{\mathrm{off}}^{(\mathrm{sim})} = 0.130\,\mathrm{s}^{-1}$
Simulation time step	$\Delta t = \{30, 10, 3, 1, 0.3\} \times 10^{-6}\,\mathrm{s}$
Binding radius	$\rho_b = 4.35 \cdot 10^{-4}\,\mu\mathrm{m}$
Unbinding radius	$\rho_{ub} = 5 \cdot \rho_b$

based approach due to the astronomical number of species involved when we consider the phosphorylation sites.

3.2.3 Biological system, ODE model, analytic solution

The biological model reactions in [39] [50] are as follows:

$$\mathsf{E_2} + \mathsf{E_2} \rightleftharpoons \mathsf{D_{22}} \;\; ; \;\; \mathsf{E_2} + \mathsf{E_3} \rightleftharpoons \mathsf{D_{23}} \;\; ; \;\; \mathsf{E_3} + \mathsf{E_3} \rightleftharpoons \mathsf{D_{33}} \tag{3.6}$$

The complete system (3.6) is too difficult to solve by hand, but with a simplification we arrive at the following reaction: $2M \rightleftharpoons D$. Using Mass Action Rate Laws we obtain the following equations of motion:

$$\frac{dM}{dt} = -k_{\mathrm{on}}M^2 + 2k_{\mathrm{off}}D \;\; ; \;\; \frac{dD}{dt} = \frac{1}{2}k_{\mathrm{on}}M^2 - k_{\mathrm{off}}D \tag{3.7}$$

This simple system has solution:

$$y(t) = \frac{(R - S_2)S_1 - (R - S_1)S_2 e^{-k_{\mathrm{on}}(S_1 - S_2)t}}{(R - S_2) - (R - S_1)e^{-k_{\mathrm{on}}(S_1 - S_2)t}} \tag{3.8}$$

The values of S_1 and S_2 are given in formula (3.9) and for more details see (4.4).

3.2.4 Extracting dimer and monomer lifetimes

We extracted the dimer and monomer lifetimes in two ways the first was by sampling the system at discrete time intervals. At each discrete time step, we checked each to see if each dimer pair was still partnered with its previous partner, and if so it was assumed that the pair did not break up; the lifetime

associated to this dimer instance would be the time step multiplied by the number of steps the dimer was observed. This method led to an inaccurate count the reason for this will be discussed in greater detail later. The error in dimer and monomer lifetimes can however be corrected by recording the time of each dimerization or dissociation event instead of observing the system at discrete time intervals.

3.2.5 Fitting the ODE solution to stochastic simulation outputs

Equilibrium

The two easiest methods for calculating the kinetic coefficients of the well-mixed model are to use either the rise via the time course of the Smoldyn algorithm or the equilibrium. When using the equilibrium, $K_D = \frac{k_{\text{on}}}{k_{\text{off}}}$ is used with fixed k_{off}.

Monte Carlo optimization

When we chose both the k_{on} and k_{off} we use the Metropolis-Hastings Algorithm, which randomly but somewhat directed selects values for both parameters until an optimum is found.

Choice of comparison points

We also have to choose our points for comparison well and the more points chosen the better the results.

3.3 Results

3.3.1 Difficulties in matching the time course of a spatial simulation with a non-spatial model

Our study was motivated by the problem of finding the proper correspondence between a spatial simulation and a traditional rate based description of the process. Comparing predictions on the time course of molecular species in a model system such as the ErbB2-ErbB3 system [39] [50], we generally found that non-spatial simulations resulted in similar, but seldom identical

behavior. We set out to investigate these discrepancies in a simple system, a reversible homo-dimerization process, represented as $2M \rightleftharpoons D$; we compare predictions of a spatial simulation similar to the one in [39][50] with the traditional ODE-based, mass-action kinetics and non-spatial stochastic simulations. We focus on the evolution of the number of molecules of monomers and dimers $M(t), D(t)$ in this system, from an initial condition of no dimers (i.e. $([M](0), [D](0)) = (R_T, 0)$), until it reaches its steady state $([\bar{M}], [\bar{D}])$.

The issues we encounter are illustrated in Figure 3.1 and 3.2. Traditionally, the dimerization-dissociation process is characterized by the forward and reverse rate constants $k_{\mathrm{on}}, k_{\mathrm{off}}$; given the initial amount of each species, the time evolution of the number (concentration) of monomers $M(t)$ and dimers $D(t)$ can be obtained as an algebraic expression (3.9),

$$M(t) = \frac{M_0(S_1 - \phi(t)S_2) - S_1 S_2(1 - \phi(t))}{M_0(1 - \phi(t)) + S_1 - S_2} \text{ where}: \ \phi(t) = e^{-k_{\mathrm{off}}\sqrt{1+4R/K_D} \cdot t}$$
$$S_1 = \frac{1}{2}K_D\left(\sqrt{1 + 4R/K_D} - 1\right) \ ; \ S_2 = -\frac{1}{2}K_D\left(\sqrt{1 + 4R/K_D} + 1\right)$$

$$(3.9)$$

(Here, M_0 is the initial amount of monomers and $R_T = M + 2D$ the (conserved) total number of receptors; $K_\mathrm{D} = k_{\mathrm{off}}/k_{\mathrm{on}}$ is the equilibrium constant; refer to the appendix 4.4 for a derivation and the origin of S_1, S_2)

A spatial simulation (following the Smoldyn algorithm) of the reversible dimerization process is characterized by six parameters: the diffusion coefficients of the two species (monomers and dimers) $D_{\mathrm{mon}}, D_{\mathrm{dim}}$, a binding and an unbinding radius (ρ_{ub}, ρ_b), a dissociation rate $k_{\mathrm{off}}^{(\mathrm{sim})}$, and a simulation time step Δt. In the simulations discussed here, we set the dimer diffusion coefficient to $1/2$ of the monomer one: $D_0 = D_{\mathrm{mon}} = 2 \cdot D_{\mathrm{dim}}$, and (unless otherwise specified), the unbinding radius is set to 5 times the binding radius, following [3] : $\rho_{ub} = 5\rho_b$. The spatial simulation provides a time course of the number of particles of each species. Due to its stochastic nature, each run of the spatial version results in a different particle number time course, as illustrated by the green lines in Fig.3.1 (A). Individual time courses, or an average over a number of runs, can be readily compared to the solution 3.9 or to non-spatial stochastic simulations.

Our goal is to identify the $(k_{\mathrm{on}}^{(\mathrm{eff})}, k_{\mathrm{off}}^{(\mathrm{eff})})$ pair for an ODE model that best approximates the average particle number time course resulting from the spatial simulation with given parameters. The most straightforward approach
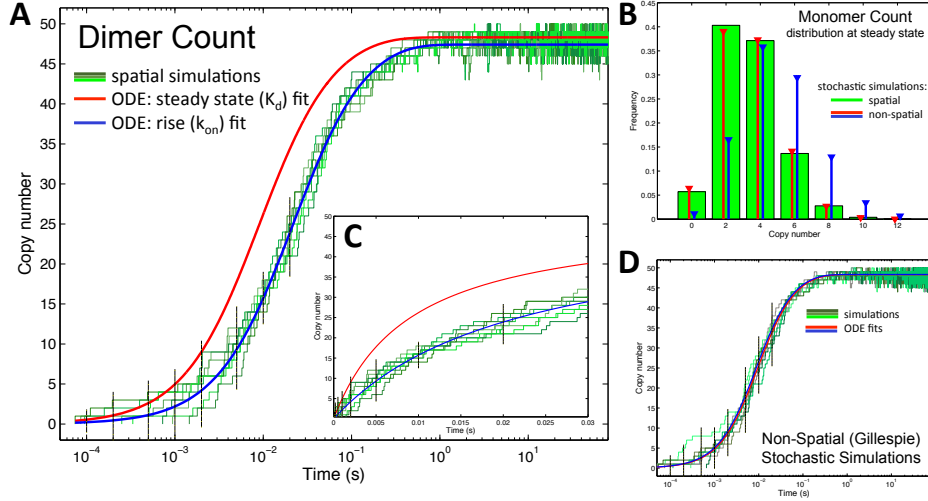
Figure 3.1: The effective well-mixed on-rate $k_{\mathrm{on}}^{(\mathrm{eff})}$ for a given spatial simulation is not uniquely defined. **A.** Time course of the number of dimers from 8 spatial simulations (green/brown) and best matching ODE time courses obtained using the same off rate $k_{\mathrm{off}}^{(\mathrm{eff})} = k_{\mathrm{off}}^{(\mathrm{sim})}$ as the spatial simulation and the on-rate $k_{\mathrm{on}}^{(\mathrm{eff})}$ (i.) determined from the mean values of the monomer and dimer counts at equilibrium (red) or (ii.) fit to the initial rise of the dimer count (blue). **B.** Steady state fitting: distribution of the number of monomers in the simulation and from well mixed stochastic (Gillespie) simulations using the parameters of theoretical curves ($k_{\mathrm{off}}^{(\mathrm{eff})} = k_{\mathrm{off}}^{(\mathrm{sim})}$ everywhere; red - $k_{\mathrm{on}}^{(\mathrm{eff})}$ from steady state, blue - rise); the steady state based prediction is very close to the spatial simulation; there is a significant discrepancy with the rise based prediction. **C.** Detail of fitting the on-rate $k_{\mathrm{on}}^{(\mathrm{eff})}$ from the rise of the number of dimers: we compare the average dimer count over the 8 spatial simulations at 8 sample time points, from $t = 10^{-4}$ s to $2 \cdot 10^{-2}$ s (from the approximate time of the first reaction to about 50% dimerization); the $k_{\mathrm{on}}^{(\mathrm{eff})}$ value minimizes the square differences, scaled by the standard deviation at each sample time. The resulting ODE curve is shown in blue; the steady state fit curve is shown in red. **D.** The procedure in (A) applied to a well-mixed stochastic (Gillespie) simulation (k_{on} matching the steady state fit to the spatial simulation); the extracted parameters are not identical, but the difference is significantly smaller, below the stochastic variability of the simulations.
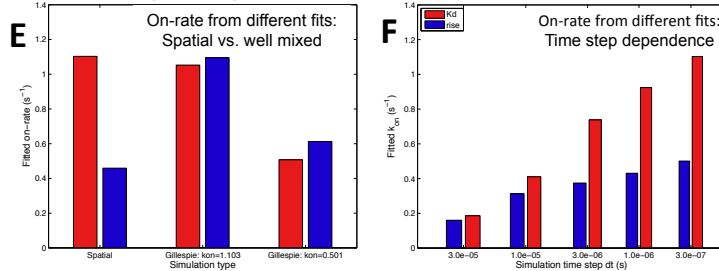
Figure 3.2: Time course fitting issues (continued). **E.** Comparison of the two $k_{\text{on}}^{(\text{eff})}$ values for the spatial simulation and two sets of Gillespie simulations. The discrepancy is significantly smaller for the non-spatial model. **F.** The two $k_{\text{on}}^{(\text{eff})}$ values for spatial simulations using different time steps. The discrepancy *increases* as $\Delta t \to 0$.

would be to identify the off-rates $k_{\text{off}}^{(\text{eff})} = k_{\text{off}}^{(\text{sim})}$ in the two models, and to estimate the on-rate $k_{\text{on}}^{(\text{eff})}$ from the simulation. There are two simple ways to do this, based on the initial behavior of the dimer count, or from the dimer / monomer ratio at steady state.

We performed several spatial simulations with several different simulation parameters, in sets of 8 individual runs for each combination of parameters. The spatial simulations shown in Figure 3.1 were performed using the "base" parameter set: dimer off-rate $k_{\text{off}}^{(\text{sim})} = 0.130\,\text{s}^{-1}$, monomer diffusion coefficient $D_{\text{mon}} = 3.1 \cdot 10^{-2}\,\frac{\text{m}^2}{\text{s}}$; binding radius $\rho_b = 4.35 \cdot 10^{-4}\,\mu\text{m}$; with an initial number of $N_{\text{rec}} = 100$ monomers (and no dimers). We set the dimer diffusion coefficient to $D_{\text{dim}} = D_{\text{mon}}/2$ and unbinding radius $\rho_{ub} = 5\rho_b$, and kept the same simulation area of $0.35\,\mu\text{m} \times 0.57\,\mu\text{m} = 0.1995\,\mu\text{m}^2$ in this set and all other simulations discussed here. Sets of 8 simulations with these parameter values were done using several time step values, $\Delta t = \{3 \cdot 10^{-5}, 10^{-5}, 3 \cdot 10^{-6}, 10^{-6}, 3 \cdot 10^{-7}\}$ s; results shown in Figure 3.1 had the smallest of these, $\Delta t = 3 \cdot 10^{-7}\,\text{s}$.

The five parameters of the spatial simulation $k_{\text{off}}^{(\text{sim})}, \rho_b, D_{\text{mon}}, N_{\text{rec}}, \Delta t$ need to be compared or mapped to the three parameters that define non-spatial simulations; $k_{\text{on}}, k_{\text{off}}, N_{\text{rec}}$. These are the same for the analytical ODE solution (the concentration is proportional to the particle copy number). For a meaningful comparison, we require the number of particles and the area to be the same in the spatial and well-mixed models; the most delicate element is the correspondence between the on-rate constant k_{on} on one hand and the

binding radius and diffusion coefficient on the other; the time step, as well as the unbinding radius and the dimer diffusion coefficient might also have an impact.

Our initial approach was to obtain an effective $k_{on}^{(eff)}$ for a given spatial simulation with its complete set of parameters. Figure 3.1 (A) shows the eight spatial simulations in green. The on-rate k_{on} was fitted to the spatial model in the two ways discussed below, resulting in two different values of the parameter. The red and blue curves represent the two different ODE solutions (3.9), using the same $k_{off} = 0.130\,\mathrm{s}^{-1}$ and $N_{rec} = 100$ as the spatial model. The red curve uses the $k_{on}^{(eff)} = 1.10294\,\mathrm{s}^{-1}$ value resulting from fitting $K_D \equiv k_{off}/k_{on}$ from the steady state concentration (copy number) of monomers $[\bar{M}]$ and dimers $[\bar{D}]$,

$$2M \xrightleftharpoons[k_{off}]{k_{on}} D \ , \quad k_{on} \cdot [\bar{M}]^2 = k_{off} \cdot [\bar{D}] \ \rightarrow \ [\bar{M}]^2 = K_D \cdot [\bar{D}] \ . \quad (3.10)$$

The blue curve in Fig.3.1(A) uses $k_{on}^{(eff)} = 0.50081\,\mathrm{s}^{-1}$, that best fits the initial rise of the number of dimers in the initially all monomer system (detail in Fig.3.1(C)).

At a first glance, the discrepancy in the steady state values of the dimer copy number $[\bar{D}]$ for the two analytical solutions in Fig.3.1(A) does not appear to be very significant. The difference is more important when comparing the steady-state number of monomers $[\bar{M}]$, which is also necessary in estimating the equilibrium constant (3.10).

As a check, we performed non-spatial stochastic (referred to as "well-mixed" or Gillespie) simulations using the same parameters and initial condition ($k_{on} = 0.130\,\mathrm{s}^{-1}$, $N_{rec} = 100$, $M(0) = N_{rec}, D(0) = 0$) with the two different $k_{on}^{(eff)}$ values for the on-rate. Fig.3.1(B) shows histograms of the number of monomers at steady state in the spatial simulations (green) and the two Gillespie simulation sets in red ($k_{on} = 1.10294\,\mathrm{s}^{-1}$) and blue ($k_{on} = 0.50081\,\mathrm{s}^{-1}$). This indicates clearerly that the lower k_{on} value is not consistent with the steady state.

Applying the same two fitting procedures to the *non-spatial simulations* results in significantly less discrepancy between the two fit modes: for a set of 8 Gillespie simulations with $k_{on}^{(sim)} = 1.103\,\mathrm{s}^{-1}$ we obtain $k_{on}^{(eff)} = 1.0525\,\mathrm{s}^{-1}$ at steady state and $k_{on}^{(eff)} = 1.4153\,\mathrm{s}^{-1}$ using the rise (Fig.3.1 (D) - compare with (A)); simulations with $k_{on}^{(sim)} = 0.5008\,\mathrm{s}^{-1}$ result in $k_{on}^{(eff)} = 0.507\,\mathrm{s}^{-1}$

and $k_{\mathrm{on}}^{(\mathrm{eff})} = 0.613\,\mathrm{s}^{-1}$ respectively. The different $k_{\mathrm{on}}^{(\mathrm{eff})}$ values are compared in Fig.3.2(E), indicating that statistical variability is unlikely to account for the discrepancy observed for the spatial simulations.[3]

Another possibly relevant factor is the time step used in the spatial simulations. In the usual practice of time-discretized schemes, results are interpreted as approximations that should converge as $\Delta t \to 0$. Even though we do not pursue such a limit in this work, the time step dependence must be taken into account as a possible cause of unexpected behavior. The bar plot in Fig.3.2 (F) compares the $k_{\mathrm{on}}^{(\mathrm{eff})}$ obtained from the steady state and the rise for sets of simulations performed with several time step values Δt, ranging over two orders of magnitude. The remarkable feature is that the discrepancy between the rise and steady state *increases as the time step is decreased.* This is an indication that the mechanism has to do with events that occur on a fast time scale, and is likely not an artifact of time discretization.

In summary, Fig.3.1 (A) illustrates the issue with trying to reproduce the time course of monomer and dimer copy numbers in a system of freely diffusing particles with a well-mixed model of the same. Using the same $k_{\mathrm{off}} = 0.130\,\mathrm{s}^{-1}$ for both models, it appears as though we cannot both fit the rise (slope of the increase of dimers, the blue line) at the same time as match the equilibrium, red line, values of dimers to monomers. The discrepancy is significant (B, C), and is not seen when comparing non-spatial stochastic simulations with the same ODE model (D, E). It also does not appear to be an artifact due to time discretization in the spatial simulation (F); rather, it becomes more significant as the time step is decreased.

3.3.2   Dimer lifetimes

Figure 3.3 (A) shows the difference between the actual dimer lifetimes, shown in red, and the apparent dimer lifetimes, shown in blue and derived by splicing together successive dimer events involving the same dimer pair (a rebinding event) whose time separation is less than $1.0 \cdot 10^{-3}\,\mathrm{s}$. Fig.3.3(A) demonstrates the issues with the off-rate k_{off} by showing how much longer it appears that a dimer survives, again in blue, than it really does, shown in red. There is a difference between the two of more than 6 seconds. This difference reiterates that k_{off} needs to be adjusted between the two types of

---

[3]The difference between $k_{\mathrm{on}}^{(\mathrm{eff})}$ values obtained through different fit modes reduces further when a set of 100 non-spatial simulations are used.
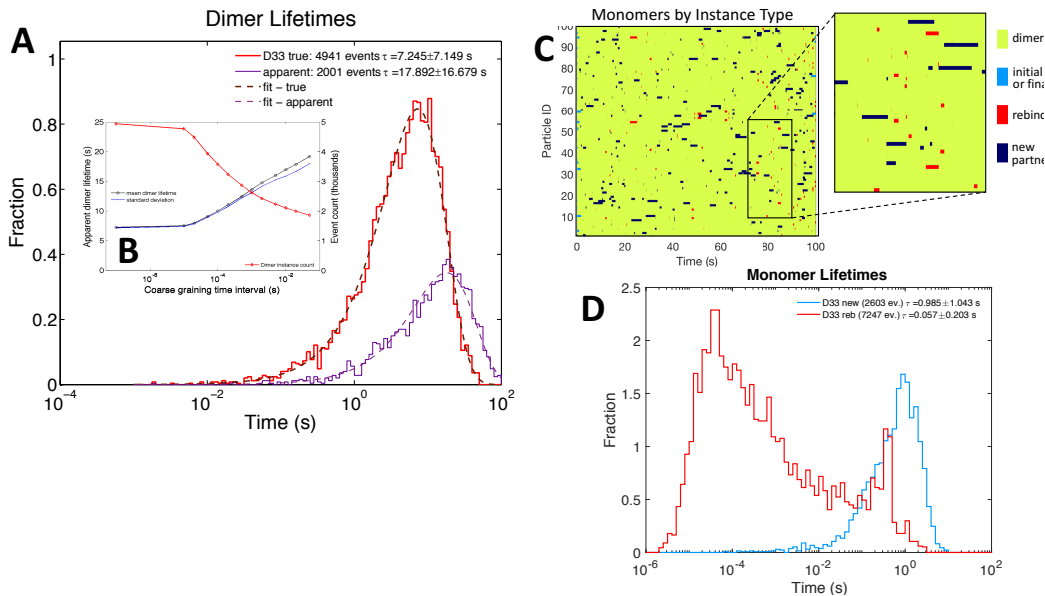
Figure 3.3: Dimer and monomer lifetimes. A. The distribution of exact dimer lifetimes (red) is consistent with the exponential distribution $f(\tau) \propto e^{-k_{\text{off}} \cdot \tau}$ (and the underlying Poisson process). To emulate the experimental method, which relies on images (snapshots) collected at time intervals $\delta T \approx 1/20\,\text{s}$, consecutive dimer events involving *the same pair* of particles separated by less than $20\,\text{s}$ were merged (*spliced* together), resulting in a smaller number of longer events. The distribution of these apparent dimer lifetimes (shown in blue / purple) is also consistent with an exponential distribution, with a larger mean. **B.** The number (red) and mean life time (blue) of the apparent dimer events obtained using different coarse graining time intervals. The true distribution is recovered only for a surprisingly small $\delta T \approx 10^{-5}\,\text{s}$. **C.** Carpet plot indicating the state (dimer or monomer) of individual receptors; dimer stretches are interrupted by very short monomer intervals. **D.** A closer examination revealed that the very short monomer events are overwhelmingly followed by dimerization *with the same partner.*

model, spatial and non-spatial, due to rebinding events. Fig.3.3 (B) deals with our choice of time intervals for the time separation when comparing apparent dimer lifetimes to actual dimer lifetimes. We see little difference between $t = 10^{-6}$ s and $t = 10^{-5}$ s . Focusing on the blue and black lines suggests we have reached an equilibrium of sorts with respect to the time step with respect to the apparent dimer lifetimes. Also the number of total dimerizing events, in red, is shown to decrease as the time step increases. In a similar fashion, there is little difference between the $t = 10^{-6}$ s and $t = 10^{-5}$ s time steps. This suggests that our simulation is best done at a time step interval of approximately $t = 10^{-5}$, and we still see the rebinding issue which confirms that we need a new k_{off} to go with the newly generated k_{on}. The non-spatial simulation sees the k_{off} differently than the spatial simulation as all particles are treated as well-mixed as soon as the dimer breaks up. The behavior of apparent dimer lifetimes becomes clearer when we look at the distribution of the time lengths of receptors spend as monomers between moments of being dimers.

Looking at Fig.3.3(C), we see that receptors tend to spend little time as a monomer as shown in blue. This tells us that if we look at time steps longer that the shortest length of time in blue we will miss count the number of dissociation events. If we look at monomer life times, seen in Fig.3.3(D), we notice that the number of rebinding events (reactions where the recently broken up monomers bind into a dimer that is indistinguishable from the previous) that involve the same pair of receptors is drastically higher than the number of partner switches and that the same pair rebindings occur much more quickly than do the partner switching events.

3.3.3 Overall fit

In the preceding subsections we found that the discrepancy between different ways of identifying the k_{on} and k_{off} parameters is caused by the presence of rebinding events, where previous dimer partners re-create the dimer over time scales that are significantly lower than the average time required for the formation of a dimer between two monomers that are placed randomly in the system. Depending on the circumstances, rebindings may be artifactual or physical; our goal here is to develop an approach to identifying the best-matching kon and koff parameters for a spatial simulation that features rebindings.

Our approach is illustrated in Fig.3 . For a given set of simulations, we
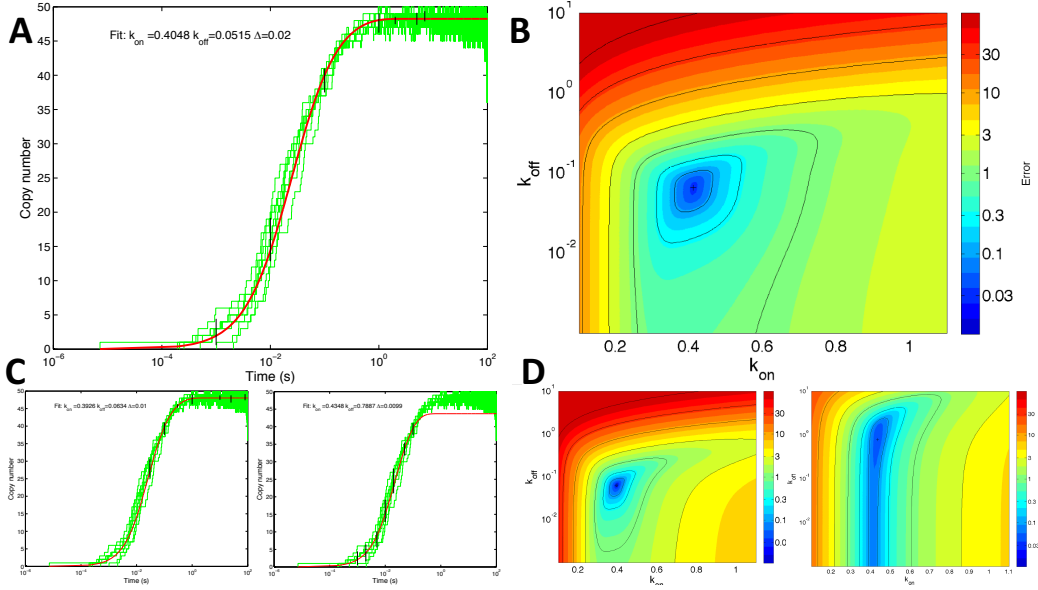
Figure 3.4: Fitting both the on-rate and the off-rate. **A.** We fit both k_{on} and k_{off} by minimizing the distance between the analytical solution (red line) and the mean of a set of simulations (green). For a set of chosen time points, we compute the difference between the mean of the simulations and the analytical expression (3.9), and divide with the standard deviation of the simulation values at the same time (black vertical bars). The sum of the squared normalized distances is the objective function to be minimized. **B.** The choice of comparison points has an influence on the optimal parameter set $k_{on}^{(eff)}, k_{off}^{(eff)}$, resulting in different fits. The two sets shown here focus on the late respectively early behavior. **C.** The optimization landscape typically has a single minimum; the landscape provides further insight into the uncertainty on the parameter values extracted and implicitly, on the quality of the fit. **D.** The landscape also varies with the choice of comparison points; (left) fitting the late behavior has uncertainty along the first diagonal (i.e. pins down the ratio k_{off}/k_{on} but less the magnitude, similarly to fitting at steady state); (right) fitting the early behavior determines k_{on} tightly, but has a large uncertainty in k_{off}, much like fitting on the rise.

consider the entire range of values for k_{on} and k_{off}; we use a measure of fit between the corresponding ODE solution and simulations in order to identify a best-match. This measure relies on a choice of comparison time points. The heat maps in Fig.3B and D represent the fitting landscape corresponding to the simulations and choices of time points in Fig.3A respectively C. We implemented a Monte-Carlo fitting algorithm that identifies the $(k_{\mathrm{on}}, k_{\mathrm{off}})$ pair corresponding to the best fit (i.e. the point with the lowest discrepancy measure) for a given set of simulations and choice of comparison points. A look at Fig.3A / B vs. Fig3C / D should make it clear that the landscape depends strongly on the choice of comparison points. .

The choice of comparison points gives us the different heat maps. We can easily see that there is a best fit choice when we choose comparison points that force a match of both the rise and the equilibrium (Fig 3 A/B). However if we match only the equilibrium or the rise we end up with a larger section of lowest distance (Fig 3 C/D). When matching the equilibrium we see a more diagonal area of blue, the lowest discrepancy (as only the ratio of k_{on} to k_{off} matter) and when matching the rise we see that the k_{off} does not matter as much as many values will still give us the lowest discrepancy again in blue. This further shows that we need to use many points for comparison in order to achieve a single pair of constant values.

Variability due to fitting method

In Fig. 3.5A we plotted the $(k_{\mathrm{on}}^{(\mathrm{eff})}, k_{\mathrm{off}}^{(\mathrm{eff})})$ pairs resulting from different fitting methods of sets of simulations performed with $D_0 = 0.031 \frac{\mathrm{m}^2}{\mathrm{s}}, N_{\mathrm{rec}} = 100$ and time steps ranging over two orders of magnitude $\Delta t = \{3 \cdot 10^{-7}, 10^{-6}, 3 \cdot 10^{-6}, 10^{-5}, 3 \cdot 10^{-5}\}$ s. The marker type indicates the fitting method, and the color indicates the simulation time step: triangles indicate a fit based only on the equilibrium ratio (K_{D}); all others represent two-parameter fits as described, using different sets of points. The sets of comparison points are indicated in the Appendix. The largest variability for the on-rate $k_{\mathrm{off}}^{(\mathrm{eff})}$, ranging over orders of magnitude, results from the two fits that focus on the initial rise in the dimer number. The variability simply reflects uncertainty in estimating the off-rate from the early dynamics of dimerization, which is dominated by dimerization events that reflect the on-rate. Conversely, the variability of the estimated on-rate $k_{\mathrm{on}}^{(\mathrm{eff})}$ is the largest in the equilibrium-based fit. The resulting values differ significantly from the corresponding ones obtained in all the two-parameter fits. The discrepancy is not random,
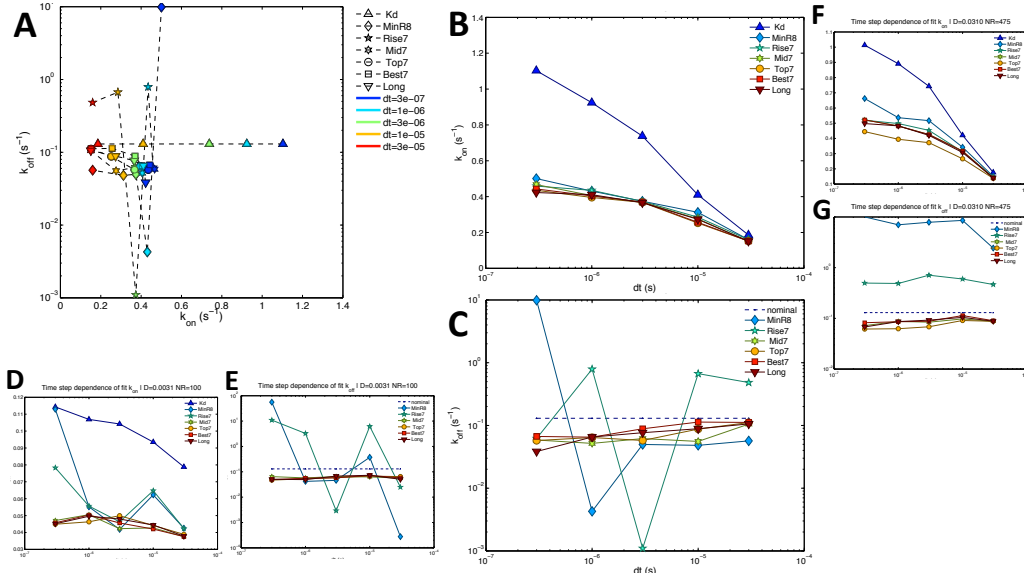
Figure 3.5: Parameter fit results across different sets of comparison
points, simulation time steps, and diffusion rate / density combina-
tions. **A.** Combined scatter plot of (k_{on}, k_{off}) fits for different com-
parison time point sets (indicated by marker shape) and simulation
time step values (indicated by color), for the "base" set of simulation
parameters $(D_0 = 0.031 \frac{m^2}{s}, N_{rec} = 100)$. **B.** On-rate values versus
simulation time step, for different comparison time sets (base simula-
tion); the discrepancy with the K_D only fit is due to the $k_{off}^{(eff)}$ value
that is different from $k_{off}^{(sim)}$. **C.** Fit of the off-rate k_{off} leads to consis-
tently smaller values than $k_{off}^{(sim)}$. The outliers are the two comparison
point sets that focus on the rise; these sets result in a very large uncer-
tainty for k_{off}. **DE.** Same as B. and C., with 10 fold reduced diffusion
coefficient $D_0 = 0.0031 \frac{m^2}{s}$. **FG.** Same as B. and C., with $N_{rec} = 475$
particles.

66

and is accompanied by a stronger dependence on the simulation time step; the discrepancy *increases* as the time step is decreased. For this, the base set of simulation parameters, the recovered on-rate $k_{\text{on}}^{\text{(eff)}}$ values are remarkably consistent *for the largest simulation time step* $\Delta t^{\text{(sim)}} = 3 \cdot 10^{-5}$ s. Other than the two outlier cases, the two-parameter fit results are fairly consistent. Still, the differences between fitting methods applied to the same exact group of simulations (especially in estimating k_{off}) are comparable to those result within the same fitting approach, for different simulation time steps.

Variability due to the simulation time step

The effective parameter values $k_{\text{on}}^{\text{(eff)}}$ and $k_{\text{off}}^{\text{(eff)}}$ obtained from different fit methods are plotted against the simulation time step $\Delta t^{\text{(sim)}}$ in Fig. 3.5, panels B and C. Fig. 3.5B reinforces the observation that the on-rate is fairly consistent between two-parameter fits, but differs significantly between those on one hand and the estimate derived from equilibrium (K_{D}) on the other. All fit methods exhibit the same trend of decreasing on-rate with increasing $\Delta t^{\text{(sim)}}$; importantly, the discrepancy *increases as* $\Delta t^{\text{(sim)}}$ *is decreased*. The behavior of the recovered off-rate $k_{\text{off}}^{\text{(eff)}}$ is less consistent. Aside the large uncertainty seen in two parameter fits focused on the initial behavior ('nominal' and 'minR8'), there is significant variability between fit modes, that is comparable to variations due to the different simulation time steps. There is no clear trend with $\Delta t^{\text{(sim)}}$, other than the recovered $k_{\text{off}}^{\text{(eff)}}$ is consistently smaller than the 'nominal' $k_{\text{off}}^{\text{(sim)}}$ and approaches it as $\Delta t^{\text{(sim)}}$ increases.

Effect of physical time / length scales

In addition to the base simulation parameters ($D_0 = 0.031 \frac{\text{m}^2}{\text{s}}, N_{\text{rec}} = 100$), we performed the same analysis (sets of simulations with varying time steps $\Delta t^{\text{(sim)}}$, followed by extraction of effective model parameters) for a physical system with the same binding radius, but reduced diffusion coefficient ($D_0 = 0.0031 \frac{\text{m}^2}{\text{s}}, N_{\text{rec}} = 100$), and for a system with the original diffusion coefficient but higher density ($D_0 = 0.031 \frac{\text{m}^2}{\text{s}}, N_{\text{rec}} = 475$).

The resulting dependence of $k_{\text{on}}^{\text{(eff)}}$ and $k_{\text{off}}^{\text{(eff)}}$ on the simulation time step is shown in Fig. 3.5 D and E, respectively panels F and G. Simulations in the 'slow' system (10-fold reduced diffusion coefficient) result in increased variability of the recovered effective parameters. The discrepancy between the

$k_{\mathrm{off}}^{(\mathrm{eff})}$ estimated at equilibrium and in the two-parameter fits (Fig. 3.5 FG) exhibits the same trends as before; however, it is larger and does not vanish even for the largest time step value. In the higher density scenario (Fig. 3.5 FG), we observe smaller variability between fit modes, and the absence of the large 'random fluctuations' seen for k_{off} estimated from two parameter fits weighted on the early behavior.

3.4   Discussion

Spatial stochastic models can connect molecular level imaging to large scale signaling models

Spatial models, from continuous reaction-diffusion to agent based simulations have been used successfully at the level of entire cells and organisms in developmental biology [46], and tumor development [16] [44]. Here we are concerned with spatial stochastic simulations of the movement and chemical state of individual receptors in the context of signal initiation. Our main motivation is the use of this type of simulation in conjunction with molecular resolution microscopy to potentially improve quantitative models of cell signaling [48] [28].

Compared to the more commonly used non-spatial ("well-mixed"), stochastic or deterministic approaches to cell signaling, spatial simulations generally provide a richer and potentially more realistic picture. This is the case especially for membrane bound receptors whose movement may be influenced by the membrane landscape [39] . In spite of mature methods and available software [51] [54], the use of spatially extended models for cell signaling is limited, due to the large added computational costs and limited benefit in terms of predictive power. The computational cost of an agent-based spatial simulation of a single dimerization-dissociation process involving N molecules is between N to N^2 times that of a well-mixed stochastic simulation of the same [25]. Models of signaling networks deal with large numbers of processes[4] [40] [8] and it is desirable to describe individual processes in terms of a small number of effective kinetic parameters.

While the generalized application of spatial simulation methods is not practical for the moment, they could be helpful in the integration of experimental data across different levels of resolution to help alleviate the problem

---

[4]there are 25 rate constants in [40]; and 85 in [8], which control $\approx 20,000$ processes

of parameter uncertainty. The predictive power of dynamical models (spatial or not) is limited by parameter uncertainty; for high dimensional systems such as signaling networks, this is exacerbated by nonlinear dynamics and the resulting possibility of qualitatively different behaviors. Particle tracking and imaging methods make individual molecular events directly accessible and could significantly improve on the parameter estimation problem. However, as our experience illustrates, the mapping between molecular level observations and cell level average behavior is not straightforward. Spatial simulations might hold the key to extracting signaling dynamics from molecular scale imaging results.

Our goal is to explore the use of spatial simulations to extract well mixed parameters (such as equilibrium constants and reaction rates) from particle tracking data. The principle is to use detailed experimental data to validate a spatial simulation that serves as a "base level model" for the experimental system; then derive a simplified model of the simulated behavior, as a controlled approximation in the mathematical sense.

The effective on-rate obtained from fitting the initial rise of the number of dimers is not consistent with the value derived from steady state

We focus on a single, reversible, homo-dimerization process, using our own implementation of a spatial simulation algorithm, most recently used in a study of ErbB2 / ErbB3 dimerization [39]. We were interested in the well-mixed rate constants k_{on}, k_{off} that best approximate the simulated time course of the number of molecules (i.e. dimers or monomers). In this algorithm, which is essentially identical to Smoldyn [3], the dimer off-rate k_{off} is also the rate at which dissociation events are triggered. We performed a set of 8 simulation runs of the dimerization-dissociation system, from an initial state with zero dimers to equilibrium (Figure 3.1A). We estimated the effective on-rate $k_{on}^{(eff)}$ by fitting the analytical solution (4.12) to time courses of the number of dimers (Fig. 3.1A and C). The resulting value $k_{on}^{(eff)} = 0.50\,\text{s}^{-1}$ was significantly different from the value $k_{on}^{(eff)} = 1.10\,\text{s}^{-1}$ that resulted from the average monomer and dimer counts at steady state (i.e. calculating K_{D} from (3.10) - Fig. 3.1A and B), using the same set of runs.

The discrepancy is specific to the spatial simulation and increases as the time step approaches zero To distinguish from effects due to the finite number of molecules, we performed two sets of 8 well-mixed stochastic (Gillespie) simulations using the $k_{\text{off}} = 0.13\,\text{s}^{-1}$ and the two different $k_{\text{on}}^{(\text{eff})}$ values obtained above. We used the two different fitting procedures on the resulting dimer count time courses. Both fitting procedures, when applied to the same group of well mixed simulations, resulted in virtually the same $k_{\text{on}}^{(\text{eff})}$ value, consistent with the on-rate that was used in the respective simulations (Figure 3.2E). Thus, the discrepancy was specific to the spatial simulation. We also performed a series of spatial simulations with different time steps (Δt) to assess the impact of the time step size. Generally, a discrete-time simulation *converges* to the physical model as $\Delta t \to 0$. In our algorithm, time discretization results in missed collisions; as the time step is reduced, the effective on-rate should increase, approaching an upper bound[5] [3]. We found (Figure 3.2F) that the discrepancy *increased* as $\Delta t \to 0$; the on-rate fitted to the initial rise in the dimer count seemed to converge, but the one resulting from steady state kept increasing. While somewhat puzzling, this indicates that the discrepancy was not a result of time discretization - or at least, not the result of an excessively large time step.

Apparent vs. true dimer lifetimes and the impact of geminate recombination

In early versions of our simulation, the complete state of the system was recorded at time intervals $\delta T \approx 0.05$ consistent with the *frame rate* $1/\delta T$ used in the experimental setup, and thus much larger than the simulation time step Δt. For each recorded configuration, dimers were identified based on the recorded (binding) state of each simulated receptor; the length or lifetime of each dimer instance was estimated based on the number of consecutive frames where that specific pair was bound. This recapitulates the way dimers are detected experimentally. It was realized early on that the *apparent dimer lifetimes* resulting from this procedure were not consistent with the expectation, the inverse of the dissociation rate $k_{\text{off}}^{(\text{sim})}$. Our simulation code records binding and unbinding events with their exact time. The distribution of the resulting *true dimer lifetimes* follows the appropriate (exponential) distribution $f(\tau) \propto \exp(-k_{\text{off}} \cdot \tau)$, where k_{off} is the off-rate used

---

[5]that corresponds to the physical Smoluchowski model

in the simulation, and provides a consistency check.

The mechanism for the significantly longer lifetimes obtained with frame-rate sampling is that a dimer observed at time T_{obs} breaks up into a pair of receptors, but the same receptors re-bind *before* the next recording time: $T_{\text{obs}} < t_{\text{off}} < t_{\text{on}} < T_{\text{obs}} + \delta T$. In the simulations discussed here, we identified all uninterrupted dimer instances in a given simulation, and sorted them by the identity of the participating receptors. To emulate sampling at a given time interval δT, we generated a "coarse grained" set of dimer events by merging consecutive dimer instances that involve the same pair of receptors, and which are separated by a time as monomers less than δT. The resulting distribution of apparent dimer lifetimes is well approximated by an exponential distribution with a longer characteristic time, consistent with a lower *apparent off-rate* (Fig. 3.3A). When performing this analysis using different coarse graining time intervals δT (Fig. 3.3B), we found that the difference between the true and apparent lifetimes decreases with δT, but the difference persists for surprisingly small values of $\delta T \approx 10^{-4}s$, several orders of magnitude smaller than the dimer lifetime.

This prompted us to look at the lengths of time spent by individual receptors as monomers, either from the beginning of the simulation to the first dimerization, or between release from a dimer (dissociation) and the next dimerization event. We further distinguished the monomer instances intervening between dimerizations, depending on the identity of the previous and next dimer partners. In Fig. 3.3C we mapped the state of the 100 receptors in one simulation; we labeled monomer instances that ended in a new partner and those that ended in rebinding with the same partner with different colors (dark blue and red, respectively). While the overall appearance of the map suggests that, in terms of the time spent as monomers, instances followed by the formation of new dimers dominate, a closer look reveals the existence of many short-lived instances of separation followed by rebinding to the same partner. The distribution of the lengths of the corresponding *monomer lifetimes* (Fig. 3.3C) reveals what in hindsight seems obvious: our simulations had a significant number of rebinding events, when the same pair of receptors dissociates, then quickly recombines. This phenomenon is called *geminate recombination* [3] and is a concern in the development of spatial simulation algorithms, such as the one we used.

Geminate recombinations: are they all bad?

Spatial stochastic simulation algorithms (SpSSA), and especially Smoldyn [3], can potentially generate artifactual geminate recombination events. The core feature of any SpSSA is that they have a representation of individual molecules and their position, and that binary reactions (typically association reactions, $A + B \rightarrow C$) are triggered by physical proximity between two specific partners, as opposed to triggering by a Poisson process and randomly selecting a pair of partners from the entire system (which is what non-spatial stochastic simulations do). This element is necessary for the ability to couple the movement of molecules and their interaction. In the Smoldyn algorithm, reactions are triggered when the potential partners are within a distance called *binding radius*, ρ_B; this quantity is a simulation parameter that characterizes each type of process.

Conversely, when a dissociation ($C \rightarrow A' + B'$) occurs, the resulting molecules necessarily emerge at a small distance from each other, consistent with physical reality; any SpSSA must have specific rules for placing the dissociation products. For reversible processes, there is a possibility of re-association; in a simulation, depending on the algorithm, the newly separated particles will emerge in positions that make recombination likely. In the Smoldyn algorithm, the distance where dissociation products are placed is a simulation parameter called the *unbinding radius* ρ_U. If $\rho_U = \rho_B$, the just-separated pair would immediately react in the next iteration; we followed the recommended practice [3] of setting $\rho_U = 5\rho_B$. The need to set an unbinding radius might be avoided in a more sophisticated approach, for example by triggering association reactions with a probability that depends on the mutual distance, or taking into account the distance as well as the relative orientation of the molecules, or any level of integrating the well-developed machinery of collision theory from particle physics [49]. However, it is not clear how such an approach might eliminate *geminate recombinations*, and indeed, whether this phenomenon is entirely unphysical.

There is no obvious, easily quantifiable, physical reason why two receptors that previously formed a dimer should *never* recombine, with or without other intervening reactions. On the other hand, it is likely that the molecular dynamics of the dissociation results in conditions (relative momentum and kinetic energy liberated by the reaction) that preclude immediate re-binding at the rate it would occur in Smoldyn if we set $\rho_U = \rho_B$. Without a more detailed knowledge of the molecular dynamics of the dimerization-dissociation

process, it is not possible to distinguish the physically correct instances of recombination from artifactual ones.

The question is *when is it unphysical* when in a SpSSA, a molecule (receptor monomer) preferentially binds to its most recent partner, to the detriment of all the other potential partners in the system. The following discussion raises questions aimed at quantifying the notion of "preference" in the selection of dimer partners.

In a well-mixed situation, where there are many available monomers, geminate recombination should be rare. Consider a pair of receptors $\{R_j, R_k\}$ that have just separated, and focus on R_j and its subsequent dimerization[s]. The probability that the next partner will be R_k (as opposed to any other receptor), should be $1/N_{\text{partners}}$, where N_{partners} represents the number of receptor monomers that R_j could possibly dimerize with. Depending on the semantics, this might be the total number of receptors in the system N_{rec}, or just the number of monomers in the system [M] immediately following the dissociation event. In the simulations discussed in Figure 3.1, we have $N_{\text{rec}} = 100$ and $[M] \approx 3$ in the spatial simulation ($[M] \approx 5$ in the well-mixed version). The "fair" probability for recombination following a dissociation event ranges from 1% to $\approx 33\%$. On one hand, it is unreasonable to accuse R_j of "bias" against receptors that are bound inside other dimers and thus unable to react with it; on the other, receptors may become available sometime after the $\{R_j, R_k\}$ breakup event. For a rational comparison, one should include a time consideration - possible partners should refer to those distinct receptors that are monomers at any time between the dissociation of the $\{R_j, R_k\}$ pair and the subsequent dimerization of R_j.

For a specific spatial simulation algorithm (in our case, Smoldyn) with a given set of parameters (diffusion coefficients, binding and unbinding radii, simulation time step), there are two aspects to consider:

Extraction of a non-spatial approximate kinetics that emulates the time course of the number of particles of each species. How does the specific spatial simulation algorithm and its parameters relate to the biological reality and particle tracking observations?

When we look only at the time course of the copy numbers of monomers and dimers, and ignore the spatial distribution of those receptors, geminate recombination (rebindings of the same receptors) appears to be the major cause of discrepancies between the spatial and well-mixed models. As we saw a great deal of discrepancies between k_{off} and k_{on} that were obtained
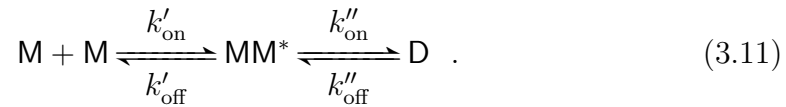
through our fitting procedures, we may be able to explain this abnormality by looking more closely at the geminate recombinations. We see from Figure 4 by comparing the base condition to both one of lower diffusion coefficient and one of high initial monomer density that the discrepancies are increased by lowering the diffusion coefficient and are decreased by increasing the initial monomer count. This is because when the diffusion coefficient is lowered the monomers have an increased difficulty finding a different partner to prevent rebindings. We also see a lowered rate of rebindings when we increased the initial number of monomers because it then becomes easier for a newly broken up dimer to find new partners for binding.

Despite setting the unbinding radius to 5 times that of the binding radius as suggested by Andrews and Bray, rebindings of former dimer partners after dissociation is an inherent feature of spatial dimerization-dissociation systems using the Smoldyn Algorithm. Some of these geminate recombination events should be regarded as simulation artifacts, however rebindings are a physical phenomenon whose importance depends on the conditions in the system. The main factor involved in these geminate recombinations is the probability that a newly dissociated monomer can find a different partner. The probability is firstly influenced by the unbinding radius which we set to 5 times that of the binding radius again as suggested by Andrews and Bray. Secondly we must consider the density of monomers as there is a proportional relationship between the density of monomers and the likelihood of finding a new partner for binding. Third, the diffusion coefficient is also a consideration in the probability that a binding is not a rebinding because the faster the diffusion the less likely a rebinding event occurs.

The geminate recombination difficulties imply that apparent dimer lifetimes will increase rapidly as the time step decreases as we miss fast dissociation-recombination events. This leads to higher apparent dimer lifetimes if the time step is too large. This issue also gives rise to a higher apparent k_{on}. Using the K_d fitting method, we see that it is higher than what is obtained by looking at the initial binding time course. These is really two sides of the same coin as if we ignore dissociations followed by recombination events altogether, it appears as though dimers dissociate less frequently, if however, we count recombination events with dimerizations involving new partners, the dimerization process appears to be faster. Regardless of the case, dimers are favored over monomers meaning that the observed K_d is smaller than purely looking at k_{on} and k_{off}.

Due to computational restrictions, most investigations of signaling net-

works do not have either the time or resources to perform spatial simulations of systems with tens to hundreds of different species. It is a critical task to derive reliable approximations (sometimes referred to as abstractions) that provide a simple description and capture the salient features of a complex phenomenon. The follow is a suggested best practice. If SPT data is available, parameter extraction should begin with identifying a spatial simulation (the set of parameters; Diffusion coefficient, binding radius, unbinding radius, and time step or etc. as required of the simulation) consistent with the experimental data. Using this and the methods previously discussed, develop the well-mixed model with k_{on} and k_{off} considering that one may have to pick a fitting method. If both steady state and early behavior are important and the discrepancy in fitting are great, we may need to introduce a new state into the system, of quasi-bound monomers MM^*,

$$\mathsf{M} + \mathsf{M} \underset{k'_{off}}{\overset{k'_{on}}{\rightleftharpoons}} \mathsf{MM}^* \underset{k''_{off}}{\overset{k''_{on}}{\rightleftharpoons}} \mathsf{D} \quad . \tag{3.11}$$

There are four parameters for this system and should be fitted to the spatial simulation. Physically, one expects the intrinsic off-rate and a measure of recombination to appear in the $\mathsf{MM}^* \rightleftharpoons \mathsf{D}$ reactions. The $\mathsf{M} + \mathsf{M} \rightleftharpoons \mathsf{MM}^*$ reactions would be slower and governed by longer effective dimer lifetimes and the apparent K_d.

Chapter 4

APPENDICES

4.1 Appendix A:Kronecker Product Explained

$$
\begin{matrix}
(\forall) & A = \{a_{ij}\} \in \mathbb{R}^{m \times n} \ , \\
\\
(\forall) & B = \{b_{kl}\} \in \mathbb{R}^{p \times q} \ ,
\end{matrix}
\qquad
A \otimes B =
\begin{pmatrix}
a_{11}B & a_{12}B & \cdots & a_{1n}B \\
a_{21}B & a_{22}B & & a_{2n}B \\
\vdots & & \ddots & \\
a_{m1}B & \cdots & & a_{mn}B
\end{pmatrix}
\quad . \quad (4.1)
$$

Note: This operation is not abelian. Also the matrices need not have any matching dimensions.

4.2 Appendix B: EGF with Microdomains

I started with a functional rules based simulation for well-mixed EGF (which has over 100 species). To add boxes, I had to take into account the perimeter and area of each box. This information had to be combined with the diffusion constant as well as information about the attractiveness of each region. I assumed that there were two types of attractive regions which could overlap. The regions differ in one tends to collect one type of EGF receptor Erb2 and the other Erb3. This difference means that the overlapping region(s) would collect heterodimers (dimers composed of both types of receptor monomers). Because of this, I treated the overlap as a different box from the other two types. This was all the additional information needed to make the simulation work for boxes.

Like the VEGF simulation in boxes, the Kron product was used to repeat the local molecular reactions, α_{loc} and β_{loc}, for each box yielding α_{mol} and β_{mol}. Also similarly to the VEGF simulation, the percent area was needed to adjust the dimerization rates inside of each box, yielding k_{mol}. Again, in a similar way to the VEGF simulation, the final step was to combine the total molecular reactions with the transfer reactions created using the previous information. This process yielded α_{total} , β_{total} and k_{total}.

I then initialized the simulation with the initial conditions being all un-phosphorlated monomers in the normal region. The results of the simulations are shown in figures. No analytical calculations were attempted for this EGF system as the number of species is well over 400.

4.3 Appendix C: Deficiency One Theorem

The Deficiency One Theorem by Feinberg in 1995 [18] proves the uniqueness of positive equilibrium for a Mass Action System of CRNs and existence of said equilibrium (for weekly reversible networks) given the conditions:

- $\delta_i \leqslant 1 \forall i = 1, \cdots, l$ where l are the linkage classes and δ_i is the deficiency of linkage class i

- $\sum_{i=1}^{l} \delta_i = \delta$ where δ is the deficiency of the CRN

- Each linkage class contains a single terminal strongly linked component

Then if the mass action system has a positive equilibrium concentration it is unique for every positive stoichiometric compatibility class and if the system is weekly reversible then moreover a positive equilibrium must exist.

4.4 Appendix D: Analytical solution for EGF homodimerization

Continuous (ODE) model

Consider a reversible homo-dimerization, with monomer species M and dimer species D.

$$2\mathsf{M} \underset{k_{\text{off}}}{\overset{k_{\text{on}}}{\rightleftharpoons}} \mathsf{D} \quad ; \quad \varphi_{2\mathsf{M}\to\mathsf{D}} = \frac{1}{2}k_{\text{on}}[\mathsf{M}]^2$$

$$\varphi_{\mathsf{D}\to 2\mathsf{M}} = \frac{1}{2}k_{\text{off}}[\mathsf{D}] \quad . \tag{4.2}$$
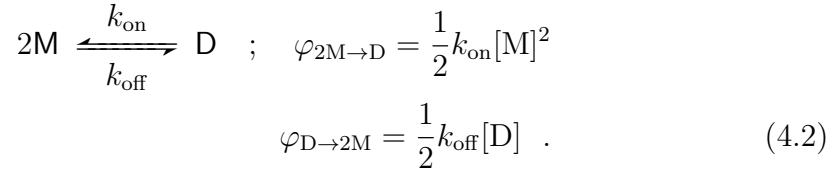
We assume mass-action rates for the fluxes φ ; k_{on} and k_{off} respectively are the physical rate constants for dimerization and dissociation are.

Traditionally, the amount of species $[\mathsf{M}], [\mathsf{D}]$ is a (molar) concentration, proportional to the number of copies of each molecule per unit volume. For the purposes of this work where we compare to simulations of individual molecules, we take the amount of each substance as the actual number of molecules in the simulation area. The physical unit for the on-rate constant k_{on} is $(s \cdot \#/\text{simspace})^{-1}$, and it is equal to the reaction probability per time for one pair of monomers in the simulation space. The off-rate constant k_{off} is simply a rate, the dissociation probability per second for one dimer.

The equations of motion are

$$\frac{d[\mathsf{M}]}{dt} = -[\mathsf{M}]^2 k_{\text{on}} + 2[\mathsf{D}]k_{\text{off}}$$

$$\frac{d[\mathsf{D}]}{dt} = \frac{1}{2}[\mathsf{M}]^2 k_{\text{on}} - [\mathsf{D}]k_{\text{off}} \quad \Rightarrow \quad [\mathsf{M}](t) + [\mathsf{D}](t) = R \tag{4.3}$$

The last equality defines the total amount of receptors R; this is conserved by the equations of motion, since $\frac{d[\mathsf{R}]}{dt} = \frac{d[\mathsf{M}]}{dt} + \frac{d[\mathsf{D}]}{dt} = 0$. The steady state condition for (4.3) is

$$2k_{\text{off}}[\mathsf{D}] = k_{\text{on}}[\mathsf{M}]^2 \Leftrightarrow 2K_{\mathsf{D}}\overline{D} = \overline{M}^2 \quad ; \quad K_{\mathsf{D}} \equiv \frac{k_{\text{off}}}{k_{\text{on}}} \quad . \tag{4.4}$$

where K_{D} is the traditional equilibrium constant.

Analytical solution

Substituting $[D] = (R - [M])/2$ in the first equation of motion, we get

$$\frac{d[M]}{dt} = k_{\text{on}}\left(-[M]^2 + K_D([M] - R)\right) = -k_{\text{on}}f([M])$$

$$f([M]) \equiv [M]^2 + K_D[M] - K_D R \quad . \tag{4.5}$$

The function $f([M])$ conveniently summarizes the equation of motion for $[M](t)$ and can also be used to defined the equilibrium condition: $f(\bar{M}) = 0$. The two roots of this polynomial are given by the quadratic formula $S_{12} = \left(-K_D \pm \sqrt{K_D^2 + 4K_D R}\right)/2$, and it is convenient to write them out as follows,

$$S_1 = \frac{K_D}{2}\left(\sqrt{1 + 4R/K_D} - 1\right) \quad ; \quad S_2 = -\frac{K_D}{2}\left(\sqrt{1 + 4R/K_D} + 1\right) \quad . \tag{4.6}$$

The two roots have the following properties (assuming $R > 0, K_D > 0$):

$$\begin{aligned}
0 &< S_1 < R & \text{because } 1 < \sqrt{1 + 4R/K_D} < 2R/K_D \\
S_2 &< 0 \, , \quad -K_D < |S_2| < K_D + R & \text{same as above} \\
S_1 &+ S_2 = K_D \, , \quad S_1 S_2 = -K_D R & \text{as roots of a quadratic}
\end{aligned} \tag{4.7}$$

Since S_1 is the only positive value of $[M]$ that verifies the condition $f([M]) = 0$, we conclude this is the only (possible) steady state, $\bar{M} = S_1$. The other solution can be written $S_2 = -(\bar{M} + K_D)$; to summarize (also using $f(x) = (x - S_1)(x - S_2)$):

$$\bar{M} = S_1 = \frac{K_D}{2}\left(\sqrt{1 + 4R/K_D} - 1\right)$$

$$f([M]) = ([M] - S_1)([M] - S_2) = ([M] - \bar{M})([M] + \bar{M} + K_D) \tag{4.8}$$

The equation of motion is a separable ODE (denote $y = [M]$):

$$y' = -k_{\text{on}}(y - S_1)(y - S_2) \rightarrow \int \frac{dy}{(y - S_1)(y - S_2)} = -\int k_{\text{on}} dt \tag{4.9}$$

rewrite the denominator under the first integral:

$$\frac{1}{(y - S_1)} - \frac{1}{(y - S_2)} = \frac{S_1 - S_2}{(y - S_1)(y - S_2)} \rightarrow$$

$$\int \left(\frac{1}{(y - S_1)} - \frac{1}{(y - S_2)}\right) dy = -\int (S_1 - S_2)k_{\text{on}} dt \tag{4.10}$$

perform the integrations and set the initial condition $y(0)$

$$\ln\left(\frac{y(t) - S_1}{y(t) - S_2}\right) = -k_{\text{on}}(S_1 - S_2)t + \mathcal{C} \ \rightarrow \ \frac{y(t) - S_1}{y(t) - S_2} = \frac{y(0) - S_1}{y(0) - S_2}e^{-k_{\text{on}}(S_1 - S_2)t}$$

$$(4.11)$$

finally, the explicit solution for $y(t) = [\text{M}](t)$ for $[\text{M}](0) = R$ works out to:

$$y(t) = \frac{(R - S_2)S_1 - (R - S_1)S_2 e^{-k_{\text{on}}(S_1 - S_2)t}}{(R - S_2) - (R - S_1)e^{-k_{\text{on}}(S_1 - S_2)t}} \quad . \tag{4.12}$$

Bibliography

[1] Eichten A, Adler AP, Cooper B, Griffith J, Wei Y, Yancopoulos GD, Lin HC & Thurston G. (2012): *Rapid decrease in tumor perfusion following VEGF blockade predicts long-term tumor growth inhibition in preclinical tumor models.* Angiogenesis 16(2), doi:10.1007/s10456-012-9328-3. Available at http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3595479.

[2] Steven S. Andrews, Nathan J. Addy, Roger Brent & Adam P. Arkin (2010): *Detailed Simulations of Cell Biology with Smoldyn 2.1.* PLoS Computation Biology 6(3), pp. 1–10, doi:doi:10.1371/journal.pcbi.1000705.

[3] Steven S Andrews & Dennis Bray (2004): *Stochastic simulation of chemical reactions with spatial resolution and single molecule detail.* Phys. Biol. 1, pp. 137–151, doi:10.1088/1478-3967/1/3/001.

[4] Aline Appert-Collin, Pierre Hubert, Gerard Cremel & Amar Bennaroune (2015): *Role of ErbB Receptors in Cancer Cell Migration and Invasion.* Frontiers in Pharmacology, doi:10.3389/fphar.2015.00283. Available at http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4657385/.

[5] Carlene Perpetua P Arceo, Joseb Editha C, Marin-Sanguinoc, Alberto & Eduardo R Mendozaa (2015): *Chemical reaction network approaches to Biochemical Systems Theory.* Mathematical Biosciences 269, pp. 135–152, doi:10.1016/j.mbs.2015.08.022. Available at https://doi.org/10.1016/j.mbs.2015.08.022.

[6] Barleon B, Totzke F, Herzog C, Blanke S, Kremmer E, Siemeister G, Marm D & Martiny-Baron G. (1997): *Mapping of the sites for ligand*

binding and receptor dimerization at the extracellular domain of the vascular endothelial growth factor receptor FLT-1. The Journal of Biological Chemistry 272(16). Available at http://www.jbc.org/content/272/16/10382.long.

[7] Hoier B, C Prats, K Qvortrup, H Pilegaard, Bangsbo J & Hellsten Y. (2013): Subcellular localization and mechanism of secretion of vascular endothelial growth factor in human skeletal muscle. FASEB Journal, doi:10.1096/fj.12-224618. Available at http://www.fasebj.org/content/early/2013/05/24/fj.12-224618.long.

[8] Dipak Barua, William S Hlavacek & Tomasz Lipniacki (2012): A Computational Model for Early Events in B Cell Antigen Receptor Signaling: Analysis of the Roles of Lyn and Fyn. J. Immunol. 189, pp. 646–658, doi:doi:10.4049/jimmunol.1102003.

[9] Yan Caia, Shixiong Xua, Jie Wuc & Quan Longb (2011): Coupled modelling of tumour angiogenesis, tumour growth and blood perfusion. Journal of Theoretical Biology 279(1), doi:10.1016/j.jtbi.2011.02.017. Available at http://www.sciencedirect.com/science/article/pii/S0022519311001196.

[10] Perrine Capmas & Herv Fernandez (2015): Effectiveness of gefitinib in combination with methotrexate in the treatment of ectopic pregnancy. Int J Womens Health, doi:10.2147/IJWH.S55556. Available at http://www.ncbi.nlm.nih.gov/pubmed/26170723.

[11] Su-Ren Chen, Xiao-Xia Hao, Yan Zhang, Shou-Long Deng, Zhi-Peng Wang, Yu-Qian Wang, Xiu-Xia Wang & Yi-Xun Liu (2016): Androgen receptor in Sertoli cells regulates DNA double-strand break repair and chromosal synapsis of spermatocytes partially through intercellular EGF-EGFR signaling. Oncotarget, Advance Publications, doi:10.18632/oncotarget.7916. Available at http://www.impactjournals.com/oncotarget/index.php?journal=oncotarget&page=article&op=view&path[]=7916&pubmed-linkout=1.

[12] Y. Chen, C Short, A.M. Halasz & J.S. Edwards (2013): The impact of high density receptor clusters on VEGF signaling. In: Second International Workshop on Hybrid Systems and Biology (HSB 2013), Electronic

Proceedings in Theoretical Computer Science 125, Taormina, pp. 84–91, doi:DOI: 10.4204/EPTCS.125.6.

[13] Ann E Cowan, Ion I Moraru, James C Schaff, Boris M Slepchenko & Leslie M Loew (2012): *Spatial modeling of cell signaling networks. Methods Cell Biol* 110, pp. 195–221.

[14] Ni CS, Sun BC, Dong XY, Sun T, Zhao N, Liu YR & Gu Q. (2013): *Promoting melanoma growth and metastasis by enhancing VEGF expression. Contemporay Oncology* 16(6), doi:10.5114/wo.2012.32486. Available at http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3687460.

[15] Peter S. Dahlberg, Blake A.Jacobson, GaneshDahal, James M.Fink, Robert A.Kratzke, Michael A.Maddaus & Lance J.Ferrin (2004): *ERBB2 Amplifications in Esophageal Adenocarcinoma. The Annals of Thoracic Surgery* 78(5), pp. 1790–1800, doi:10.1016/j.athoracsur.2004.05.037. Available at https://doi.org/10.1016/j.athoracsur.2004.05.037.

[16] Thomas S Deisboeck, Zhihui Wang, Paul Macklin & Vittorio Cristini (2011): *Multiscale Cancer Modeling. Annu. Rev. Biomed. Eng.* 13, pp. 127–155.

[17] Radek Erban & S Jonathan Chapman (2009): *Stochastic modelling of reaction-diffusion processes: algorithms for bimolecular reactions. Phys. Biol.* 6, p. 046001.

[18] M. Feinberg (1995): *The existence and uniqueness of steady states for a class of chemical reaction networks. Archive for Rational Mechanics and Analysis* 132, doi:10.1007/BF00375614. Available at https://doi.org/10.1007/BF00375614.

[19] Martin Feinberg (1987): *Chemical reaction network structure and the stability of complex isothermal reactorsI. The deficiency zero and deficiency one theorems. Chemical Engineering Science* 42(10), doi:10.1016/0009-2509(87)80099-4. Available at https://doi.org/10.1016/0009-2509(87)80099-4.

[20] Kathryn M. Ferguson (2008): *A structure-based view of Epidermal Growth Factor Receptor regulation. Annu Rev Biophys* 37,

doi:10.1146/annurev.biophys.37.032807.125829. Available at http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2745238/.

[21] Feilim Mac Gabhann & Aleksander S. Popel (2007): *Dimerization of VEGF receptors and implications for signal transduction: A computational study.* Biophysical Chemistry 128(2-3), doi:10.1016/j.bpc.2007.03.010. Available at http://www.sciencedirect.com/science/article/pii/S0301462207000798.

[22] Daniel T. Gillespie (1977): *Exact Stochastic Simulation of Coupled Chemical Reactions.* The Journal of Physical Chemistry 81(25), doi:10.1021/j100540a008. Available at http://pubs.acs.org/doi/pdf/10.1021/j100540a008.

[23] David Z. Goodson (2011): *Mathematical Methods for Physical and Analytical Chemisty.* Wiley.

[24] A. M. Halasz, H. J. Lai, M. M. Pryor, K Rdhakrishnan & J.S. Edwards (2013): *Analytical Solution of Steady-State Equations for Chemical Reaction Networks with Bilinear Rate Laws.* IEEE Trans. Comp. Biol. Bioinf. 10(4), pp. 957–969.

[25] Ádám Miklós Halász, Meghan McCabe Pryor, Bridget S. Wilson & Jeremy S. Edwards (2017): *Spatiotemporal Modeling of Membrane Receptors.* In F. Graw, F. Matth:aus & J. Pahle, editors: *Modeling Cellular Systems, Contributions in Mathematical and Computational Sciences* 11, Springer International Publishing, Switzerland, pp. 1–37.

[26] Martin Hermann & Masoud Saravi (2014): *A First Course in Ordinary Differential Equations.* Springer, New Delhi.

[27] William S. Hlavacek, James R. Faeder, Michael L. Blinov, Alan S. Perelson & Byron Goldstein (2003): *The Complexity of Complexes in Signal Transduction.* The American Physical Society 84(7), doi:10.1002/bit.10842. Available at http://www.t6.lanl.gov/wish/documents/BiotechnolBioengin03.pdf.

[28] A Holcman, N Hoze & Z Schuss (2015): *Analysis and Interpretation of Superresolution Single-Particle Trajectories.* Biophys. J 109, pp. 1761–1771.

[29] Se-Hoon Hong, Won Jae Lee, Young Doo Kim, Hyunjoo Kim, Young-Jun Jeon, Bitna Lim, Dong-Hyung Cho, Won Do Heo, Doo-Hyun Yang, Chan-Young Kim, Han-Kwang Yang, Jin Kuk Yang & Yong-Keun Jung (2016): *APIP, an ERBB3-binding partner, stimulates erbB2-3 heterodimer formation to promote tumorigenesis. Oncotarget, Advance Publications*, doi:10.18632/oncotarget.7802. Available at http://www.impactjournals.com/oncotarget/index.php?journal=oncotarget&page=article&op=view&path[]=7802&pubmed-linkout=1.

[30] Tsung-I Hsu, Yi-Chang Want, Chia-Yang Hung, Chun-Hui Yu, Wu-Chou Su, Wen-Chang Chang & Jan-Jong Hung (2016): *Positive feedback regulation between IL10 and EGFR promotes lung cancer formation. Oncotarget, Advance Publications*, doi:10.18632/oncotarget.7894. Available at http://www.impactjournals.com/oncotarget/index.php?journal=oncotarget&page=article&op=view&path[]=7894&pubmed-linkout=1.

[31] Chen Huang, Yongsheng Xu, Xuemin Li & Wei Wang (2013): *Vascular endothelial growth factor A polymorphisms and age-related macular degeneration: a systematic review and meta-analysis. Mol Vis.* Available at http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3675058/.

[32] Pablo A Iglesias & Brian P Ingalls (2010): *Control Theory and Systems Biology.* MIT Press.

[33] Olfert IM, Howlett RA, Wagner PD & Breen EC. (2010): *Myocyte vascular endothelial growth factor is required for exercise-induced skeletal muscle angiogenesis. American Journal of Physiology Regulatory, Integrative and Comparative Physiology* 299(4), doi:10.1152/ajpregu.00347.2010. Available at http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2957383/.

[34] Md. Rashedul Islam, Kazuki Yamagami, Yuka Yoshii & Nobuhiko Yamauchi (2016): *Growth factor induced proliferation, migration, and lumen formation of rat endometrial epithelial cells in vitro. Journal of Reproduction and Development*, doi:10.1262/jrd.2015-158. Available at https://www.jstage.jst.go.jp/article/jrd/advpub/0/advpub_2015-158/_article.

[35] Merenmies J, Parada LF & Henkemeyer M. (1997): *Receptor tyrosine kinase signaling in vascular development. Cell Growth and Differentiation: the molecular biology journal of the American Association for Cancer Research* 8(1). Available at http://cgd.aacrjournals.org/cgi/reprint/8/1/3.

[36] Saxton M J (1995): *Single-particle tracking: effects of corrals. Biophysical Journal.* Available at http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1236263/pdf/biophysj00058-0093.pdf.

[37] Faeder James (2011): *Toward a comprehensive language for biological systems. BMC Biology* 9(68), doi:10.1186/1741-7007-9-68. Available at http://www.biomedcentral.com/1741-7007/9/68.

[38] Merav Kedmi, Aldema Sas-Chen & Yosef Yarden (2015): *MicroRNAs and Growth Factors: An Alliance Propelling Tumor Progression. J Clin Med.*, doi:10.3390/jcm4081578. Available at http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4555078/.

[39] R Kerketta, Ádám M Halász, Mara P Steinkamp, Bridget S Wilson & Jeremy S Edwards (2016): *Effect of Spatial Inhomogeneities on the Membrane Surface on Receptor Dimerization and Signal Initiation. Front. Cell Dev. Biol.* 4(81), pp. 1–13.

[40] Boris N. Kholodenko, Oleg V. Demin, Gisela Moehren & Jan B. Hoek (1999): *Quantification of Short Term Signaling by the Epidermal Growth Factor Receptor. The Journal of Biological Chemistry*, doi:10.1074/jbc.274.42.30169. Available at http://www.ncbi.nlm.nih.gov/pubmed/10514507.

[41] Michael Klann, Arnab Ganguly & Heinz Koeppl (2012): *Hybrid spatial Gillespie and particle tracking simulation. Bioinformatics* 28(ECCB2012), pp. i549–i555.

[42] Tsoi M, Lague MN, Boyer A, Paquet M, Nadeau ME & Boerboom D. (2013): *Anti-VEGFA Therapy Reduces Tumor Growth and Extends Survival in a Murine Model of Ovarian Granulosa Cell Tumor. Translational Oncology* 6(3). Available at http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3660791.

[43] L. Claesson-Welsh M. Simons, E. Gordon (2016): *Mechanisms and regulation of endothelial VEGF receptor signalling.* Nat Rev Mol Cell Biol., doi:10.1038/nrm.2016.87. Available at http://www.ncbi.nlm.nih.gov/pubmed/27461391.

[44] Paul Macklin, Hermann B Frieboes, Jessica L Sparks, Ahmadreza Ghaffarizadeh, Samuel H Friedman, Edwin F Juarez, Edmond Jonckheere & Shannon M Mumenthaler (2016): *Systems Biology of Tumor Microenvironment*, chapter Progress Towards Computational 3-D Multicellular Systems Biology, pp. 225–246. *Advances in Experimental Medicine and Biology* 936, Springer International Publishing, Switzerland.

[45] Meghan M. McCabe, Shalini T. Low-Nam, Adam Halasz, Diane S. Lidke, Bridget S. Wilson & Jeremy S. Edwards (2013): *Dynamic transition states of ErbB1 phosphorylation predicted by spatial-stochastic modeling.* In preparation.

[46] H G Othmer, K Painter, D umulis & C Xue (2009): *The intersection of theory and application in elucidating pattern formation in developmental biology. Math. Model. Nat. Phenom.* 4(4), pp. 3–82.

[47] Martine Perrot-Applanat & Melanie Di Benedetto (2012): *Autocrine functions of VEGF in breast tumor cells Adhesion, survival, migration and invasion. Cell Adh Migr.*, doi:10.4161/cam.23332. Available at http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3547902/.

[48] Fredrik Persson, MArtin Linden, Cecilia Unoson & Johan Elf (2013): *Extracting intracellular diffusive states and transition rates from single-molecule tracking data.* Nat. Meth. 10(3), pp. 265–271. Doi:10.1038/nmeth.2367.

[49] Alexander V Popov & Noam Agmon (2001): *Three-dimensional simulations of reversible bimolecular interactions: The simple target problem. J. Chem. Phys.* 115(9), pp. 8921–8932.

[50] Meghan McCabe Pryor, Mara P Steinkamp, Ádám M Halász, Ye Chen, Shujie Yang, Marilyn S Smith, George Zahoransky-Kohalmi, Mark Swift, Xiao-Ping Xu, Dorit Hanien, Niels Volkmann, Diane S Lidke, Jeremy S Edwards & Bridget S Wilson (2015): *Orchestration of ErbB3*

signaling through heterointeractions and homointeractions. MBoC 26, pp. 4109–4123.

[51] Martin Robinson, Steven S. Andrews & Radek Erban (2015): *Multiscale reaction-diffusion simulations with Smoldyn.* Bioinformatics 31(14), pp. 2406–2408.

[52] Guo-Xiang Ruan & Andrius Kazlauskas (2012): *Axl is essential for VEGF-A-dependent activation of PI3K/Akt.* The EMBO Journal, doi:10.1038/emboj.2012.21. Available at http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3321201/.

[53] Michael J. Saxton & Ken Jacobson (1997): *SINGLE-PARTICLE TRACKING:Applications to Membrane Dynamics.* Annual Review of Biophysics and Biomolecular Structure 26(1), pp. 373–399, doi:10.1146/annurev.biophys.26.1.373. Available at https://doi.org/10.1146/annurev.biophys.26.1.373. PMID: 9241424.

[54] Melanie I Stefan, Thomas M Bartol, Terrence J Sejnowksi & Mary B Kennedy (2014): *Multi-state Modeling of Biomolecules.* PLOS Comp. Biol. 10(9), p. e1003844.

[55] Alda Tufro & Delma Veron (2012): *VEGF AND PODOCYTES IN DIABETIC NEPHROPATHY.* Semin Nephrol., doi:10.1016/j.semnephrol.2012.06.010. Available at http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3438453/.

[56] Prakash Vempati, Aleksander S. Popel & Feilim Mac Gabhann (2013): *Extracellular regulation of VEGF: isoforms, proteolysis, and vascular patterning.* Cytokine Growth Factor Rev., doi:10.1016/j.cytogfr.2013.11.002. Available at http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3977708/.

[57] Kallirroi Voudouri, Aikaterini Berdiaki, Maria Tzardi, George N. Tzanakakis & Dragana Nikitovic (2015): *Insulin-Like Growth Factor and Epidermal Growth Factor Signaling in Breast Cancer Cell Growth: Focus on Endocrine Resistant Disease.* Anal Cell Pathol (Amst)., doi:10.1155/2015/975495. Available at http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4518167/.

[58] Zhixiang Wang (2016): *Transactivation of Epidermal Growth Factor Receptor by G Protein-Coupled Receptors: Recent Progress, Challenges and Future Research. Int J Mol Sci*, doi:10.3390/ijms17010095. Available at http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4730337/.

[59] Argraves WS, Larue AC, Fleming PA & Drake CJ. (2002): *VEGF signaling is required for the assembly but not the maintenance of embryonic blood vessels. Developmental Dynamics* 225(3), doi:10.1002/dvdy.10162. Available at http://onlinelibrary.wiley.com/doi/10.1002/dvdy.10162/abstract;jsessionid=5BBFFD183D77C1C5F205A18627FAAE07.d01t01.

[60] M.X. Sliwkowski Y. Yarden (2001): *Untangling the ErbB signalling network. Nat Rev Mol Cell Biol.*, doi:10.1038/35052073. Available at http://www.ncbi.nlm.nih.gov/pubmed/11252954.

[61] Jin Yang, Michael I. Monine, James R. Faeder & William S. Hlavacek (2008): *Kinetic Monte Carlo method for rule-based modeling of biochemical networks. The American Physical Society*, doi:10.1103/PhysRevE.78.031910. Available at http://www.ccbb.pitt.edu/Faculty/Faeder/Publications/Reprints/Yang2008.pdf.

[62] Gur Pines Yosef Yarden (2012): *The ERBB network: at last, cancer therapy meets systems biology. Nature Reviews Cancer* 12, doi:10.1038/nrc3309. Available at http://www.nature.com/nrc/journal/v12/n8/full/nrc3309.html.

[63] J. Zhang, K. Leiderman, J. Pfeiffer, B. Wilson, J. Oliver & S. Steinberg (2005): *Characterizing the Topography of Membrane Receptors and Signaling Molecules from Spatial Patterns Obtained using Nanometer-scale Electron-dense Probes and Electron Microscopy. Micron* 37, doi:10.1016/j.micron.2005.03.014. Available at http://www.math.unm.edu/~stanly/prints/GoldStats.pdf.