New Methods in Computational Systems Biology

A Thesis

Submitted to the Faculty

of

Drexel University

by

David J. Miller

in partial fulfillment of the

requirements for the degree

of

Doctor of Philosophy

March 2008

Dedications

For my wife, Jaime Wolfe.

Acknowledgments

The author thanks his advisor Avijit Ghosh, along with the current and former members of the Computational Biophysics Group: Dhruv Pant, Ray Zou, Tom Shortell, Yihua Wang, Hanbing Lin, Aparna Kumar, Travis Hoppe, and Geoff Lukas.

Portions of this work were supported in part by a NSF Bioinformatics Starter Grant and by a grant from the Pennsylvania Tobacco Health Funds Initiative.

Table of Contents

List of Tables

List of Figures

Abstract

New Methods in Computational Systems Biology
David J. Miller
Avijit Ghosh, Ph.D.

Systems biology strives to reach greater understanding of biological function through an integrative, multidisciplinary approach utilizing experimentation, theory, and simulation in equal measures. Drawing from the traditionally distinct fields of biology, chemistry, physics, engineering, mathematics, computer science, informatics, and medicine, systems biology regards biological components as acting in tandem in a unified hierarchical system over a wide range of scales, from nano-scale (proteins and small molecules) to micro-scale (organelles and cells) to macro-scale (tissue and organs).

Within this burgeoning field, computational modeling of cell signaling serves not only to validate theoretical and experimental findings, but also to provide quantitative and even predictive analysis of biochemical networks and intracellular machinery.

In this thesis, a model of the canonical MAPK signal transduction pathway (well studied for its role in a large percentage of cancers) is analyzed using the custom simulation software package `CellSim` as a tool for predicting targets for effective anti-cancer drugs, as well as predicting the effects of such drugs on non-cancerous cells. Furthermore, computational tools and methods are developed for extending such purely kinetic models of intracellular signaling into the spatio-temporal realm, introducing locality, transport, and cell geometry.

Chapter 1: A Systems Biology Approach to Oncogene Detection

1.1 Summary

Computational models of signal transduction pathways have been successful both qualitatively as well as quantitatively in describing how complex protein networks control cell function. Moreover, the study of networks has been used to elucidate not only how these pathways control the complex regulation and response mechanism of cells, but also provide insight into how a breakdown in the biological circuitry can lead to particular disease states.

We have recently examined the circuitry within the MAPK signal transduction pathway to understand how changes within this canonical network may lead to malfunction, notably the rise of proto-oncogenic cells. In addition we have developed a new complementary technique that provides insight into which key players within the pathway are most likely to be most conducive to selective inhibition within this transformed line of cells. These tools have been made freely available to the public, as part of a software suite developed by our group, `CellSim` [2], and an overview will be given on how `CellSim` may be used to quantitate cell function and, moreover, malfunction.

1.2 Introduction

Computational Biophysics has always played a complementary role to the experimental biological sciences. The role of a computational biophysicist, as such, is not to develop tools that simply reassure the experimentalists that well-regarded experiments may, in fact, be duplicated *in silico*, but instead must also provide new predictive and quantitative tools that provide new insight into biological mechanism or function. New tools from the development of new experimentally designed united-residue force fields such as UNRES [3] to new special purpose hardware techniques such as

MDGRAPE [4] have given rise to new predictive mathematical and computational methods that can probe behavior of individual proteins on a femtosecond scale. It is in this sense that new tools such as systems biology have been developed to address the complementary issue of how these proteins can act *en masse* to dictate not only form, but intra-cellular behavior [5, 6, 7, 8, 9]. The scale is perhaps different (from angstrom to micrometer; from picosecond to millisecond; from individual proteins to micromolar concentrations), but the philosophy is the same. The method of systems biology is to take experimental data that is relatively simple to reproduce, and to use it to provide insight into phenomena that cannot be easily elucidated by an existing set of experiments.

Systems biology uses these experimentally derived, computational techniques to demonstrate both how cells have the ability to respond to external stimuli as well as how intra-cellular signaling circuitry essential for cell function is controlled. The mechanism by which this occurs depends fundamentally on the way in which cells use protein networks as the mechanism for translating extra-cellular signals into intra-cellular behavior. Hence, the complexity of signal transduction networks is based on the interplay between different aspects of the signaling process, any of which may change with subtle external or internal changes to the cell [1]. The focus of our group is on one particularly important signaling cascade: the MAPK signal transduction pathway.

1.3 The canonical MAPK signal transduction pathway

The canonical MAPK signaling cascade (Figure 1.1) is one of the most well-studied signal cascades, both experimentally and computationally [10]. This central cascade is critical for governing cell growth and proliferation as well as actin cytoskeleton rearrangement [11, 12, 13, 14]. Stimulation of the cascade activates many downstream effectors including PI3K [15], Bcl-2 [16], and PKC [17, 18], among others [10]. The central cascade is activated via the following mechanism: The epidermal growth factor (EGF) signaling begins with the epidermal growth factor receptor (EGFR) and traverses a series of signaling proteins to the Ras protein. The Ras protein works in part by activating a series of kinases starting from Raf (a Mitogen Activated Protein Kinase Kinase Kinase (MAPKKK)), which activates the Mitogen Activated Protein ERK Kinase (MEK). This in turn activates the extracellular-signal regulation kinase (ERK), which subsequently translocates to the nucleus and

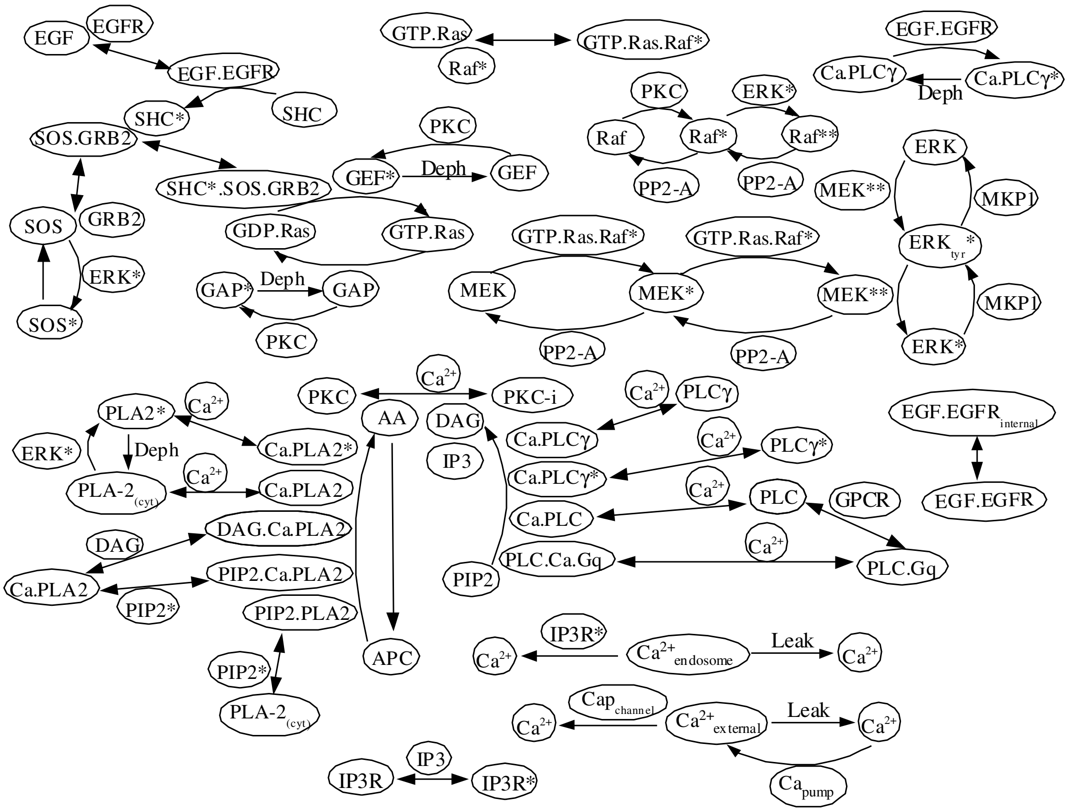


Figure 1.1: The canonical MAPK pathway

stimulates a series of growth promoting transcription factors [19]. The pathways described represent a simplified description of the full process of cell signaling, as this cascade is but a single member of a complex set of parallel, interacting pathways [1].

1.3.1 Oncogenic transformation of the MAPK pathway

The central member of the MAPK pathway, Ras, illustrates the importance of transformations within the MAPK signal transduction cascade. The Ras protein is a GTP-binding signaling protein, activating downstream effectors by binding to GTP, while inactive in the GDP bound form. The Ras oncogene was first found experimentally by its ability to induce tumor-like growth in fibroblasts [20, 21, 22]. The major oncogenic transformation involves specific mutations in Ras that prevent hydrolysis of GTP-bound Ras by GTPase Activating Proteins (GAP), leaving Ras continuously activated (and thus persistently signaling downstream effectors). These particular mutations have been implicated in approximately 30% of all human cancers, predominantly in lung, colon and pancreatic cancers [23]. This is likely an under-estimation of oncogenic transformation of Ras related cascades as mutations in other effectors downstream may cause a similar transformation in absence of mutations in the Ras gene itself. The transformation to tumor cells does not occur by a mutational event in Ras alone, but through a series of malignant transformations along or between several distinct pathway species [24].

1.4 Methods

Systems biology methods developed by our group use either ordinary differential equations (ODEs) or partial differential equations (PDEs) to describe the overall temporal or spatio-temporal behavior of the protein network within (and between) compartments of a cell [8, 9]. Enzymatic reactions and other chemical interactions are represented as simply a system of ODEs that couple to active and passive transport. Passive transport includes processes such as simple diffusive processes. Active transport includes explicit advective terms, modeling transport along actin filaments and other ATP driven processes.

Elementary chemical reactions describe the enzymatic and non-enzymatic reactions within each compartment. These reactions may be written as:

$$\sum_i n_i R_i \underset{k_b}{\overset{k_f}{\rightleftharpoons}} \sum_j n_j P_j \tag{1.1}$$

where a set of reactant species R_i with stoichiometric coefficients n_i inter-converts into a set of product species P_j with stoichiometric coefficients n_j and rate constants k_f and k_b. Due to the fact that chemical collisions greater than bimolecular are rare, the order of an elementary chemical reaction is not typically greater than two. Characteristic of signaling pathways are enzymatic reactions such as phosphorylation or dephosphorylation events. These reactions may be expressed as a combination of a reversible and an irreversible chemical reaction as follows:

$$E + S \underset{k_2}{\overset{k_1}{\rightleftharpoons}} E.S \xrightarrow{k_3} E + S^* \tag{1.2}$$

where E represents an enzyme which catalyzes the substrate S. The intermediate species $E.S$ first forms reversibly with rate constants k_1 and k_2, followed by an irreversible catalytic step with rate constant k_3, which releases the activated substrate S* and the enzyme for further catalysis.

These reactions lead to a set of ordinary differential equations such that one may express the time rate of change in concentration of all species as a system of unimolecular and bimolecular reactions such that:

$$\frac{d[C_i]}{dt} = \sum_j k_{ij}[C_j] + \sum_{l>m} k_{ilm}[C_l][C_m] + \sum_j T(C_i, C_j) \tag{1.3}$$

where k_{ij} is the rate constant for a unimolecular reaction involving species C_i and C_j at concentrations $[C_i]$ and $[C_j]$ respectively. If k_{ilm} is positive, then k_{ilm} represents the rate constant of formation of species C_i from a bimolecular reaction between species C_l and C_m with concentrations $[C_l]$ and $[C_m]$. Conversely, if k_{ilm} is negative, then k_{ilm} represents the rate constant of disassociation of species C_i into two species, C_l and C_m. $T(C_i, C_j)$ represents a function governing the passive transport of a species C_i into a different compartment via passive channels, after which it is labeled with a subscript j as C_j .

1.4.1 Mutations

A mutation in a particular gene in a signaling pathway manifests itself in one of two ways. In the first case, the mutation may directly affect the *interaction* between two species. If two species A and B reversibly associate/disassociate with rate constants k_f/k_b:

$$A + B \underset{k_b}{\overset{k_f}{\rightleftharpoons}} A.B \tag{1.4}$$

then a mutation of this type will perturb k_f or k_b by some amount. For instance, lowering k_f by some amount represents a mutation that hinders the ability of species A or B to associate into $A.B$. We define a mutation of this type as an “interaction” mutation. This mutation may occur in either species A or B, as the effect is the same. A simple analysis using Arrhenius theory may be used to connect the free energy change from a mutation with the corresponding kinetic parameters k_f and k_b [7, 6].

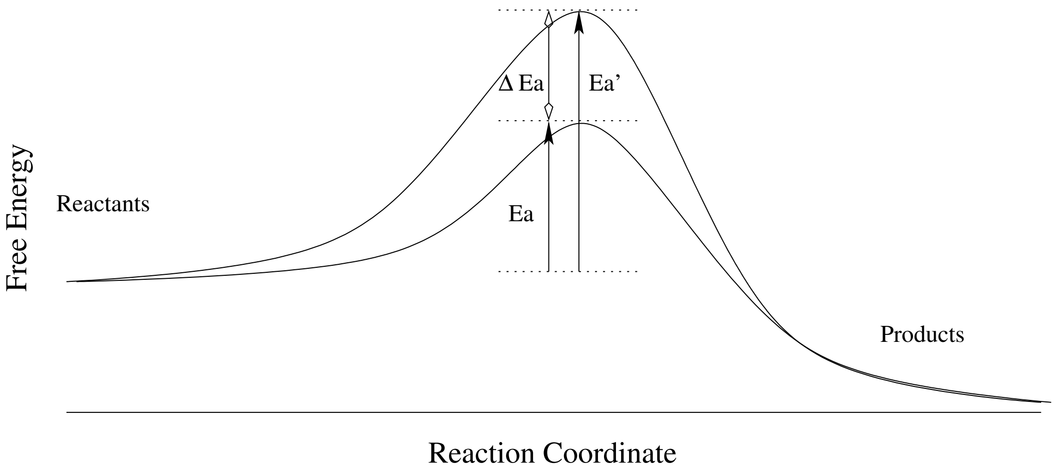


Figure 1.2: Graphical representation of the energy barriers involved in a reaction. An increase in the barrier height lowers the rate constant. The lower rate constant k'_f corresponds to the greater barrier height $E^{'}_a$. The effect of a mutation changes the free energy along the reaction coordinate by an amount ΔE, leading to a change in the forward kinetic rate by an amount proportional to $exp\{-\Delta E/kT\}$

The rate of interconversion from reactant to product may be given as:

$$k_f = Ce^{-E_a/k_BT} \tag{1.5}$$

where C is a constant prefactor, E_a is the barrier energy of activation and k_B is Boltzmann's constant. Mutations in the enzyme may affect a transition rate by either increasing the barrier height or changing the free energy of the initial state of the system by some amount ΔE. The new mutated system may therefore be considered to be a perturbed system with a new barrier of height E'_a and a forward transition rate of k'_f (Figure 1.4.1). The ratio k'_f/k_f is :

$$k'_f/k_f = \frac{e^{-(E_a+\Delta E)/k_bT}}{e^{-E_a/k_bT}} \tag{1.6}$$

and simplifying

$$k'_f/k_f = e^{-\Delta E/k_bT} \tag{1.7}$$

The right side of Equation (1.7) is a function of the *change* in the barrier height and not of the barrier height E_a itself.

Equation 1.7 can be used to simulate the effect of any single mutation on the normal MAPK signal transduction pathway without having to explicitly delineate the underlying cause. The key effect governing the transformation of the normal MAPK signal transduction pathway is the ability to activate downstream ERK without EGF stimulation. Moreover, such mutations may be rank ordered in terms of ΔE to yeld a ranked list of predicted proto-oncogenes. See [6, 7] for a full discussion of this approach.

1.5 Drug targeting

With drug development costs now reaching 500 million dollars or more, development strategies represent a significant hurdle in bringing new therapeutics to the marketplace [25]. The experimental development cycle can be optimized in a manner that minimizes the number of false positives during costly clinical trials using systems biology techniques to detect proto-oncogenes *in silico*. Rather than modeling the effect of a mutation, modeling of an inhibitor may be performed with the addition of a single chemical reaction representing simple competitive binding between the substrate and the target protein:

$$I + S \underset{k_b}{\overset{k_f}{\rightleftharpoons}} I.S \tag{1.8}$$

The binding free energy can be calculated from the equilibrium constant k_{eq} of the reversible binding reaction. A particular inhibitor may bind to any substrate within the MAPK pathway. The efficacy of the inhibitor against a particular target is gauged by its ability to stop auto-activation of the entire set of “mutated” pathways described in the previous section. Furthermore, targets that successfully inhibit all these pathways can further ranked by the minimum binding affinity and concentration required. Complementarily, effective drug targets can also be ranked by their ability to *avoid* effecting non-mutated pathways.

1.6 Concluding Remarks

Computational biophysics has been successful in underscoring how quantitation and simulation can be used to address difficult problems of interest in biology. Systems biology continues this tradition with an emphasis on the *macroscopic* rather than the *microscopic*, focusing on not single molecules or proteins but rather how entire systems interact. As illustrated, systems biology can be used to not only quantitate how behavior is governed, but additionally how malfunctions in the signaling process can give rise to aberrant signaling processes. Finally, an analysis of these aberrant networks can be used to suggest novel treatment strategies based upon both predicted treatment effectiveness and patient tolerance.

Chapter 2: A Systems Biology Approach for Lead Drug Target Prediction with Applications to the MAPK Signal Transduction Pathway

2.1 Summary

In silico models of signal transduction pathways have been highly successful in describing, quantitatively, how complex protein networks govern overall cell function. By analyzing a recently developed model of oncogenesis in the Mitogen Activated Protein Kinase (MAPK) signal transduction pathway, a quantitative ranking of putative targets that inhibit cells with mutations within the MAPK pathway has been developed. The inhibitor, a virtual drug, is constructed by specifying its parameters: initial concentration (drug dosage) and dissociation constant k_d. Many of the targets found by this analysis have inhibitors that are currently under investigation. Several novel targets not previously investigated have been found. Of the thirteen targets, Ras, Guanine Exchange Factor (GEF), and Raf, show the highest potential. In addition, the analysis finds that certain calcium blockers may have potential as anti-tumor agents, functioning at much lower concentrations but requiring higher binding affinity.

2.2 Introduction

The advantage of a systems approach in drug research is that *in silico* models can be manipulated in a more subtle way than global inhibition experiments. As such, these models are a useful method for discriminating between various ways a molecule influences a cascade of events. Computational models allow one to thoroughly and efficiently test a wide range of hypothetical inhibitors and gauge the overall pathway response inexpensively, compared to traditional experimental methods.

One of the focuses of drug design in cancer research has been the MAPK signal transduction pathway, one of the central signaling cascades in the cell [26, 27, 28, 29, 30]. The MAPK pathway mediates signaling of growth hormones and cytokines and participates in regulating both cellular proliferation and death [31, 32]. There are many mutations known to occur in this pathway that cause unregulated activation of Extracellular Regulated Kinase (ERK), which in turn leads to activation of corresponding proliferative pathways in many tumor lines [33, 34]. Mutations of components in the MAPK pathway have been detected in 48% of tumors in small cell lung cancer [35], 50% of colon cancers [36], 75% of pancreatic cancers [37] and 27% of breast cancers [38], among others. A systems biology approach to the MAPK pathway has motivated complex signaling models by several groups, including Bhalla and coworkers [1, 26], Schoeberl and coworkers [39], Kholodenko [40], Levchenko [41], and others.

The goal of this computational analysis is to predict drug targets that will stop the uncontrolled proliferation of tumor cells. Drugs that target the MAPK pathway have shown much promise in preventing further growth of such tumors, and several are currently in clinical trials [29, 30, 42]. These include Farnesyl Transferase Inhibitors [43, 44, 45], Raf inhibitors [46, 47], and MAPK-ERK Kinase (MEK) inhibitors [48], among others.

The procedure outlined is a fast, effective computational assay for lead drug targeting, and complements previous work by our group on mutations in the MAPK signal transduction pathway [7, 6]. EGF-directed stimulation of ERK in the MAPK pathway occurs through a complex series of reactions (Figure 2.1). A simplified outline of the main central pathway may be described, beginning with extracellular EGF binding to the EGF receptor on the cell membrane. This induces receptor dimerization and autophosphorylation [49]. The bound receptor forms a docking site for the signaling molecule complex of SOS-Grb2, which then activates the G-protein Ras by stimulating the exchange of Guanosine Diphosphate (GDP) with Guanosine Triphosphate (GTP). This causes a conformational change in Ras, enabling it to bind to Raf, leading to Raf activation. Activated Raf phosphorylates and activates MEK, which in turn activates ERK via phosphorylation. Phosphorylated ERK (ERK*) can then enter the nucleus and phosphorylate transcription factors, such as Elk-1 and other Ets family proteins [28, 50, 51].

This description of the central cascade fails to convey the complete pathway's inherent complexity, as the central pathway is itself controlled through several feedback and feed forward loops that can act as bistable switches within the network (Figure 2.1). These have been described in a series of seminal papers by Bhalla and coworkers [1, 52]. This central pathway interacts with the PLC-γ-PKC pathway at two critical points: Protein Kinase C (PKC) activates both Raf and Ras, which in turn leads to the activation of ERK in a positive feedback loop, and ERK* activates Phospholipase A2 (PLA2), which results in the activation of PKC via diacylglycerol (DAG), leading to another coupled feedback loop. Conversely, ERK* down-regulates its own activation through the double phosphorylation of Raf in a negative feedback loop [53]. In addition, ERK* also phosphorylates Son of Sevenless (SOS) leading to the inactivation of signaling from the receptor in another negative feedback loop. A careful balance of positive and negative feedback loops within these coupled pathways dictate the response [54, 55]. Mutations within these cascades lead to an imbalance and manifest themselves in either uncontrolled proliferation or cell death.

As shown in previous work by our group, the characteristics of the pathway itself predict which mutations in the signal cascade will give rise to proliferation in a normally functioning cell [7, 6]. These changes can be classified into two types: 1) mutations that lead to a change in an interaction between that species and one of its interacting partners and 2) mutations that lead to a change in concentration of that species within the cell. An example of the first type may be a change in an active site that causes a particular protein-protein or protein-ligand interaction to be hindered or completely blocked. The effect of this type of mutation leads to changes in the free energy along the reaction coordinate of each of the relevant biological reactions in which the mutated protein participates. Illustrative of this are the oncogenic Ras mutations that prevent Ras-GAP interactions or prevent GTP hydrolysis [56]. The key point is that the effect of a mutation is to affect progression along the reaction coordinate adversely. This implies that a mutation in a particular substrate manifests itself in a temporal model of the signaling pathway by affecting the chemical kinetics of association or disassociation of that substrate and its interacting partner or of the irreversible catalytic step. The effect of these mutations is a change in the free energy along the reaction coordinate by some amount ΔE (Figure 1.4.1) [7]. As an example of the second type, mutations may also be classified as those that lead to an over-expression or under-expression of a particular enzyme either directly or indirectly. The canonical example of this is the over-expression of the EGF Receptor in breast cancer tumors [38].

The effect of these types of changes in a normal signal transduction pathway can be correlated with changes in the underlying kinetics or initial conditions of a complex protein network. Top ranked mutations that are predicted to give rise to activated ERK levels are listed in Tables 2.1 and 2.2. The prediction is based on measuring the biochemical activity of the pathway due to a change in the free energy of a species within the pathway. These tables will be described in more detail in the next section.

2.3 Methods

In silico drugs are used to treat the generated pathways developed in our previous work with the goal of inhibiting high levels of activated ERK that characterize the mutated cells. In selecting prime targets within the pathway, two criteria are used: (1) the binding affinity and (2) the initial concentration of a hypothetical drug that interacts with that target. The most efficacious drug target is one that requires a low initial drug concentration and a low drug binding affinity to elicit the primary response. Each model drug is designed to inhibit a single species within the pathway with a predetermined binding strength and initial concentration. A desirable target is defined as a target that, when inhibited, causes the activated ERK concentration to fall to basal levels. Furthermore, of all the targets with this property, the most optimal target is the one that requires the lowest binding affinity. In general, when one designs therapeutic agents against a particular target, one looks for the drug with the highest binding affinity. When looking for targets to bind to, however, one looks for the complement. The reason for this is that the binding affinity is directly related to the free energy of binding. The free energy of binding is directly related to the set of intermolecular interactions between the ligand and substrate. These intermolecular interactions include specific hydrogen bonding patterns, electrostatic, and steric considerations. These intermolecular interactions are reduced when a ligand binds to a substrate with a lower binding affinity. This expands the scope and range of possible inhibitors that will effectively stop the activation of this cascade.

Each hypothetical drug is tested against a set of perturbed pathways that have elevated activated ERK levels described previously [7, 6]. Each row in Tables 2.1 and 2.2 represents a different mutation that leads to activated ERK levels. To illustrate, the third highest ranked mutation in Table 2.1 involves the inhibition of the interaction between Raf and Ras. This results in a lower disassociation rate by inhibiting the reverse rate constant by a multiplicative factor (k_{mod}) of 0.6. Row 3 in Table 2.1 states that any value of k_{mod} that is 0.6 and lower for this reaction will lead to activation of the MAPK pathway without stimulation. The multiplicative factor k_{mod} is proportional to the change in free energy along the reaction coordinate due to a random mutation in either Ras or Raf by:

$$\Delta E = -k_{Boltz} T ln \left(k_{mod} \right) \tag{2.1}$$

where k_{Boltz} is Boltzmann's constant and T is the temperature (Figure 1.4.1). The effect of this mutation is to change the reverse rate constant k_b from $0.5 s^{-1}$ to $0.3 s^{-1}$. This particular mutation represents one that inhibits Raf and Ras-GTP disassociation. The set of reactions listed in Table 2.1 represents 20 possible deactivating mutations that lead to unstimulated activation. The direction column labels whether k_{mod} is applied to the forward or reverse rate constant. Similarly, Table 2.2 represents an additional 20 possible activating mutations that lead to activation of the MAPK pathway. Activating mutations are defined by reactions such that $k_{mod} > 1$, leading to an increase in the corresponding rate constant. The family of mutations listed in Tables 2.1 and 2.2 represent a total of 40 virtual cell lines that lead to activated ERK levels without stimulation.

Tables 2.1 and 2.2 do not represent all possible mutations within a cell but are simply the ones that are highly ranked by measuring the biochemical activity of the MAPK pathway (Figure 2.1) via the activation of ERK. It is not expected that a cell with only a single mutation from either table will cause transformation of a real cell. These single mutations must be considered within a framework of additional mutations in pathways not modeled, which would otherwise inhibit the biochemical activity of the MAPK pathway. However, the predicted single mutations should be considered to be leading indicators of mutations that are likely when coupled to mutations in these additional mechanisms. As such, the targeting of the MAPK pathway must be considered as a single possible intervention point in the treatment of possible transformed cells. The approach outlined can be generalized quite easily, and it is expected that a similar strategy can be used to target these other pathways.

2.4 *In Silico* Model of Drug Targets

To calculate the effects of a competitive inhibitor on a particular virtual cell line, one adds a single chemical reaction representing simple binding between the substrate and the target protein:

$$P + I \underset{k_b}{\overset{k_f}{\rightleftharpoons}} P.I \tag{2.2}$$

The binding free energy can be calculated from the binding affinity k_d (k_b/k_f) of the reaction. A particular inhibitor may bind to any substrate within the MAPK pathway. The efficacy of the inhibitor against a particular target is gauged by its ability to stop auto-activation of the pathway of the entire set of cells described in the previous section. Furthermore, targets that successfully inhibit all transformed cell lines are further ranked by the minimum binding affinity and concentration needed to deactivate all cell lines. While competitive binding is of course the simplest way that a drug may interact with a particular protein, it should be noted that in real systems drugs may non-competitively bind or may engage in allosteric inhibition and so forth.

Having a high binding energy is equivalent to making the association irreversible or equivalently, setting $k_b = 0$. In the MAPK model used (Figure 2.1), there are 31 potential drug target candidates. The range of binding strengths used to test the targets range from a k_d value of $10^{-6}\mu M$ to $10^{4}\mu M$ separated by logged increments of $10^{0.1}\mu M$. As the steady state response is being investigated, k_f is set to 1 and the value of k_b is adjusted accordingly. The tested concentration range for the inhibitor lies between $1pM$ and $1\mu M$ in log increments of $10^{0.1}\mu M$. Figure 2.2 is illustrative of the results of this calculation. In Figure 2.2, a drug targets Phospholipase C-γ (PLC-γ) through a range of concentrations and the number of successfully inhibited pathways is tabulated. The contour plots outline the number of successfully inhibited pathways (Tables 2.1 and 2.2). There is a clearly delineated region where the drug at a particular concentration can inhibit all pathways (the top left portion of the figure) and a region where the drug fails to inhibit any of the pathways (bottom right). For drugs targeted against PLC-γ, a k_d value of $10^{-1}\mu M$ or lower and any concentration above $10^{-1.25}\mu M$ will be sufficient to inhibit all pathways. At higher k_d values, a higher concentration will be needed to compensate for the lower binding affinity of the target. There is a certain trade off between binding affinity and concentration. The results against all targets are shown in Figures 2.2-2.7. The figures themselves are sorted by the minimum concentration the drug needs to function against a particular target and within a particular concentration sorted by k_d. The figures have been grouped by similarity and then qualitatively ranked by efficacy.

Figure 2.2 shows the first set of targets that are classified as "high concentration" drug targets. Inhibitors against PLC, PLC-γ and G-GDP all need a concentration of approximately $10^{-1}\mu M$ to inhibit all the mutation containing pathways. G-GDP refers to the G-$\alpha\beta\gamma$ trimer associated with GDP. At that concentration, the binding affinities required are such that PLC-γ inhibitors need a slightly higher minimum binding affinity than PLC. PLC inhibitors, in turn, require a slightly higher minimum binding affinity than G-GDP. It is worth noting that PLC-γ also needs a slightly higher concentration to fully inhibit all mutated pathways, clearly demarking it as a less promising drug target than either of the other two candidates. While the numerical values are likely not to correspond with actual binding affinities or concentrations needed for inhibition of real pathways, it is this qualitative ranking that provides a simple estimate of plausible targets.

Figure 2.3 consists of only a hypothetical drug designed against the kinase ERK. Drugs targeted against ERK need a concentration of $10^{-1.5}\mu M$ to work successfully and have a similar k_d range to that of G-GDP and PLC as shown in Figure 2.2. Of the drug targets in Figures 2.2 and 2.3, a clear hierarchy of targets exists, such that ERK targeted drugs provide the best opportunity for inhibition of proliferative pathways.

The targets in Figure 2.4 reqire a drug with a minimum concentration of $10^{-1.75}\mu M$ to be successfully inhibited. PKC-cytosolic and PLA2 are the two members of this category. While PLA2 inhibitors require a relatively low concentration, they also require a high affinity to function. In contrast, PKC inhibitors will also function at similar concentrations but additionally inhibit the set of mutated pathways at a much larger k_d range than any previously described drug. The range is essentially flat up to $10^{-0.5}\mu M$ at a concentration of $10^{-1.75}\mu M$.

MEK, the only target in Figure 2.5, needs a minimum concentration of $10^{-2.0}\mu M$ to function. MEK illustrates the trade-off in choosing a particular target: while inhibitors against MEK will function at lower concentrations, a higher binding affinity is needed to inhibit all pathways. That said, the binding affinity needed is quite similar to most of the previous targets, the notable exception being PKC.

The targets in Figure 2.6 need a drug concentration of $10^{-2.5}\mu M$ to successfully inhibit the set of mutation containing pathways. The enzymes Ras, Raf, and GEF are all in this category. It is interesting to note that a minimum k_d value of $10^{-1}\mu M$ is needed for all three drug targets. In terms of k_d, these targets are similar to most of the previously described targets except that they continue to function at lower concentrations than previously described targets.

The final set of targets includes calcium regulatory proteins (Figure 2.7). These targets are markedly different from all previously described targets. These targets regulate calcium flow to the cytosol via various mechanisms. The "Inact Cap Entry" target represents endosomal capacitive entry channels that are blocked when Ca^{2+} is sequestered into endosomes. The "Capacitive Ca Entry" is the transient receptor potential (TRP) plasma membrane channel. The "Ca Leak from Extracell" target represents the pool of Ca^{2+} leak channels on the plasma membrane. These "calcium blocker" drugs work at much lower concentrations than any other previously listed drug target. "Capacitive Ca Entry" and "Inact Cap Entry" inhibiters work at concentrations up to $10^{-3.6}\mu M$ but require a high binding affinity. The maximum k_d range is approximately $10^{-4.5}\mu M$ for these targets. Similarly a hypothetical calcium leak blocker would function at concentrations up to $10^{-4.5}\mu M$ with a maximum k_d of approximately $10^{-4}\mu M$. The nature of the calcium blockers is markedly different from the targets analyzed previously and will be discussed in more detail in the next section.

2.5 Discussion

As there is a clear trade-off between the minimum required binding affinity and initial concentration, it is difficult to define a clear hierarchical scale that rates the "best" target. A summary of the results is shown in Figure 2.8. Each point in the figure shows the minimum concentration and maximum k_d

value that still inhibits a majority of the mutation-containing pathways. Many of the highly ranked targets shown are currently under experimental investigation.

Of the targets in Figure 2.2, PLC and PLC-γ inhibitors are known to have anti-cancer properties. PLC-γ inhibitor peptides have been found to block cancer cell growth in colorectal carcinoma cells (KMS-4, KMS-8 cell lines) [57, 58]. Evodonin and coworkers have recently found that the PLC inhibitor U73122 can significantly reduce the migration of human tumor cells [59]. In a related study of U73122, it was found that PLC-γ is the rate limiting step for governing tumor cell invasion in various tumor cell lines including prostate cancer (DU-145) and several breast cancer lines (MDA-468, MDA-231, MDA-361) [60]. Oh and coworkers have recently discovered a new inhibitor based on a fungal strain that inhibits PLC activity with potent cytotoxicity against breast cancer (T47D), prostate cancer (PC-3) and ovarian cancer [61]. There is some evidence that U73122 has the side-effect of releasing Ca^{2+} from the intracellular stores by directly activating ion channels [62]. This may make PLC a less than ideal target. In contrast, inhibitors of large G-proteins associated with the MAPK pathway are not currently used in cancer therapy.

Inhibitors for ERK (Figure 2.3) are currently not well studied, and only a few ERK inhibitors are currently known to exist. A small oligonucleotide made from the amino-terminal end portion of MEK has been recently shown to specifically target ERK1/2. While this peptide can inhibit ERK *in vitro*, it does not inhibit ERK *in vivo* as it cannot pass the membrane barrier [63]. The cyclin-dependent Kinase (CDK) inhibitor Purvalanol has recently been shown to not only target CDKs but ERK as well [64]. The antiproliferative activity of Purvalanol has been shown to be mediated by the inhibition of both ERK and CDK inhibition.

The next set of highly ranked targets found by our study is the PKC and PLA2 kinases. Several other PKC inhibitors are currently under investigation as possible anti-cancer agents. Koivunen and coworkers have recently tested several PKC inhibitors including Go6976, Safingol, Rottlerin and Bisindolylmaleimide I on cultured urinary bladder carcinoma cells (5647 and T24). Of this set of inhibitors, they found that Go6976 has potential as an anti-cancer drug, as it has the additional effect of inhibiting metastasis [65]. Schwartz and coworkers have found that the PKC inhibitor

SPC-100221 stops cancer cell invasion of gastric cancer [66]. The use of PKC inhibitors in cancer therapy has been reviewed by several groups [67, 68, 69]. Tamoxifen, an antiestrogen, is widely used in breast cancer treatment. Tamoxifen is also a PKC inhibitor, though its activity is in the micromolar range and the effect of PKC inhibition is likely not to play a large role in the use of Tamoxifen in treatment [70, 71, 72].

PLA2 inhibitors are currently used as anti-inflammatory agents [73, 74] and recently PLA2 has been proposed as a target for therapeutic intervention in cancer [75]. The PLA2 inhibitor Quinacrine has been shown to be effective against prostate cancer when used with the anti-tumor drug Paclitexal [76], however there exists conflicting evidence that Quinacrine use may also lead to increased cancer risk [77, 78]. At this time, no major PLA2 inhibitor is currently in use as a therapeutic agent.

Several MEK inhibitors are currently in various stages of development [48]. Astra-Zeneca is currently developing AZD6244/ARRY-142886, a MEK inhibitor that has been shown to be tumor suppressive in melanoma, pancreatic, colon, lung, and breast cancers. This drug has completed Phase I trials and is currently starting Phase II clinical trials. Pfizer is currently testing the MEK inhibitor CI-1040 in Phase I and Phase II clinical trials for breast cancer, colon cancer, and melanoma [79]. Additionally, studies of the MEK inhibitor PD98059 have shown that MEK inhibition impairs proliferation of cancer cells [80, 81]. The MEK inhibitor PD98059 has recently been shown to suppress metastasis of the prostate cancer cell lines [82]. PD98059 has also been used to enhance the efficacy of cancer treatment. For instance, it has been found to increase the success rate of Docetaxel in suppressing the proliferation of prostate cancer cell lines and increasing apoptosis. The efficacy of this combinatorial approach was shown to be far more effective than either drug alone [83].

Excluding the calcium blockers, inhibitors against Ras, Raf, and GEF are predicted by the method to be the most efficacious drug leads. Ras and Raf inhibitors are some of the most well studied drug leads within the MAPK pathway, and several candidates are currently in clinical trials. As Farnesyl Transferase is responsible for collocating Ras to the plasma membrane for activation, several Farnesyltransferase inhibitors are currently in clinical trials [84, 85, 86]. These inhibitors include R115577, which shows promise for leukemia and colorectal cancer and is currently in Phase I and Phase II clinical trials [44, 87, 88, 89, 90, 91, 92, 93]. Bristol Myers Squibb currently has shown that BMS-214462 displays cytotoxicity against human tumor cell lines [94, 95, 96, 97, 98, 99, 100, 101]. Schering Plough is currently investigating SCH-66336, currently in Phase I and Phase II clinical trials against a wide variety of human tumors [102, 103, 104, 105, 106, 107, 108, 109, 110].

There are several new Raf inhibitors that are also currently in clinical trials. Bayer Pharmaceuticals is currently testing Bay43-9006 in Phase I and II clinical trials [111, 112, 113, 114, 115, 116, 117]. Isis Pharmaceuticals is currently testing ISIS-5132, an oligonucleotide based inhibitor that functions by down-regulating the expression of Raf and is currently in Phase I and Phase II clinical trials [118, 119, 120, 121, 122, 123, 124, 125, 126]. The last target in this set, GEF, does not presently have an inhibitor in clinical trials, despite the existance of GEF inhibitors such as Brefeldin-A [127] and TRIPα [128]. Nontheless, GEF has recently garnered interest as a target for cancer therapy [129, 130, 127]. While elevated levels of GEF have been found in a variety of cancers, particulary pancreatic [131], GEF's role in cell apoptosis may also paradoxically enable it for use in therapies involving gene transfer to tumors [132] as opposed to a target of inhibitory drugs.

The final set of targets, the calcium channels (Figure 2.7), can be inhibited with drugs at extremely low concentrations, although they require a higher binding affinity than the previous drug targets. There is much conflicting data in the literature on the relationship between calcium blockers and cancer. It has been proposed that calcium blockers actually increase the risk for cancer [133, 134, 135]. This work, however, has been somewhat controversial [133, 134, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151], as more recent studies have failed to find any significant correlation between calcium blockers and cancer [152, 153, 154, 155]. Additionally, Debes et.al. found an inverse correlation between calcium blockers and prostate cancer [156], Lee et.al. found that T-type calcium channel blockers inhibit human cancer cell growth [157] and the calcium blocker Verapamil has been found to stop tumor growth in several studies [158, 159, 160, 161, 162, 163, 164]. There is also evidence that Nicardipine works in a similar fashion [165, 166, 167].

2.6 Conclusion

The method developed is an extension of previous work done by our group [7, 6]. Through direct simulation, it is shown that drug function can be clearly delineated by drug k_d and initial concentration. Thirteen kinases within the pathway are shown to have potential as anti-tumor targets, and there is considerable evidence in the literature that inhibitors against these targets have anti-tumor activity. As there is a trade-off in binding affinity and concentration, a simple ranking is overly simplistic. In general, however, targets further towards the bottom right of Figure 2.8 are preferable, and it is clear from the analysis that some drug targets are better suited. Of these, Ras, Raf, and GEF targets show the most promise, and can be inhibited by drugs at the lowest concentrations. It was also found that calcium blockers, while requiring a high binding affinity, function at very low concentrations. It is possible that certain calcium blockers may function as anti-tumor agents, perhaps in conjunction with standard therapies.

The advantage of such *in silico* testing procedures is that the analysis is straightforward both conceptually and computationally and can provide a simple, fast gauge to focus experimental efforts as well as point out new leads not previously known. The limitation of these types of systems biology approaches is that they are dependent on a reasonably complete set of kinetic parameters. As the MAPK pathway is one of the most well-studied pathways available, it is well suited for this analysis, and it is envisioned that as newer pathways become available in public databases, such analyses may become commonplace.

Several immediate *in silico* extensions to the work described herein exist. A combinatorial approach of simulating multiple inhibitors in tandem is one such extention that has a dirct experimental analogue to the use of synergistically acting drug cocktails. Another extension, outlined in the next chapter, is the simulation of these inhibitors on the wild-type MAPK pathway. This of course models the action of targeted drugs on non-cancerous cells, i.e. side effects. Finally, the introduction of locality and transport into models of the MAPK signal transduction pathway leads to an obvious spatio-temporal complement of the targeting strategy described herein[168, 8].

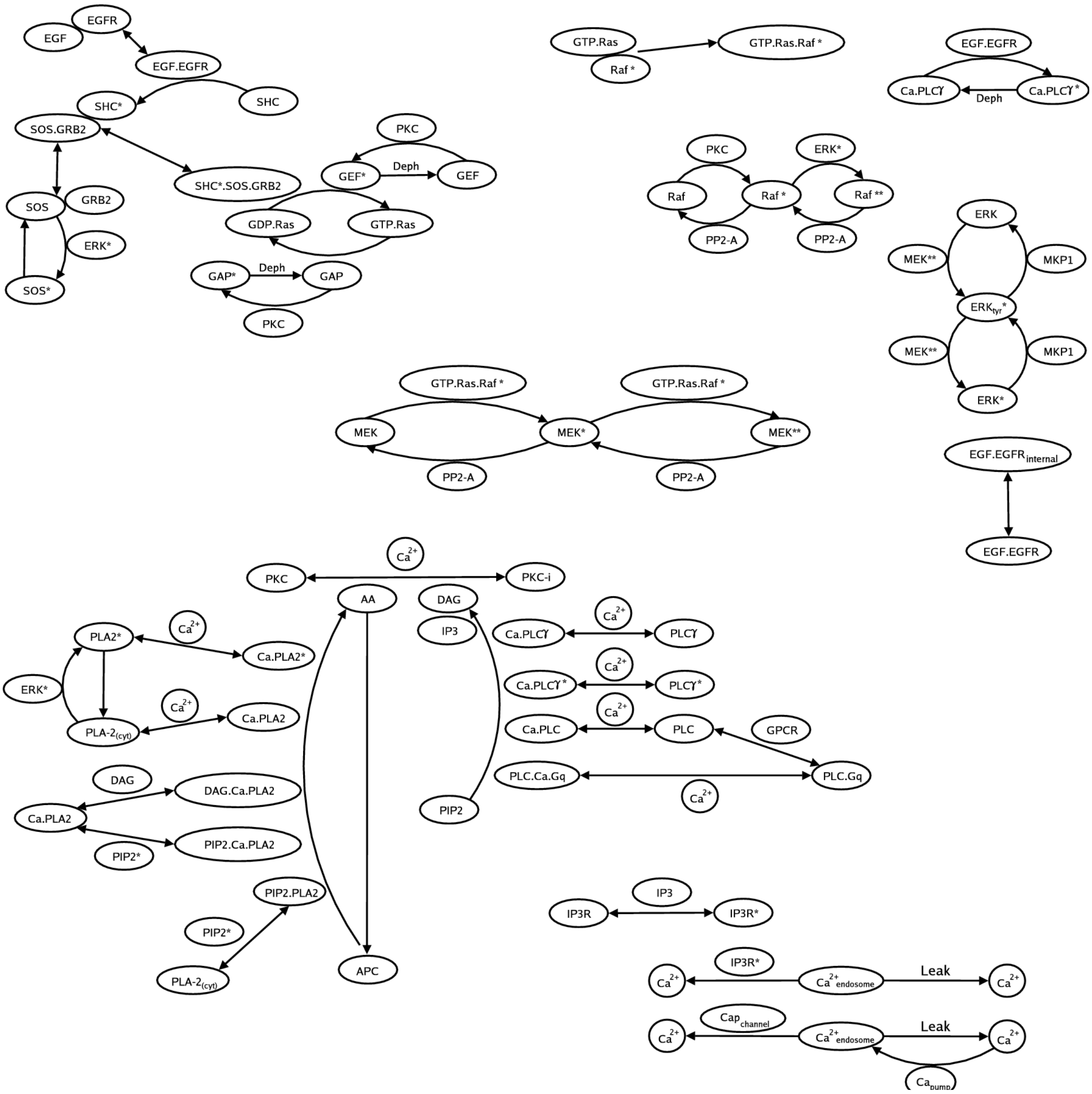


Figure 2.1: The main pathways used in this model. The arcs connect species involved in enzymatic (such as phosphorylation/dephosphorylation) reactions. For example, Phosphatidylinositol bisphosphate (PIP2) conversion to DAG and Inositol 1,4,5 Trisphosphate (IP3) is carried out by four enzyme complexes Ca-PLC, Ca-Gα-GTP-PLC, Ca-PLCγ and Ca-PLCγ^* of the two isoforms of PLC. An asterisk denotes the addition of a phosphate group on that species. The association and disassociation reactions are shown by straight arrows. The parameters for this model are taken from the work of Bhalla and Iyengar [1], where a full discussion of the naming conventions, signalling feedback loops, and individual parameters used may be found.

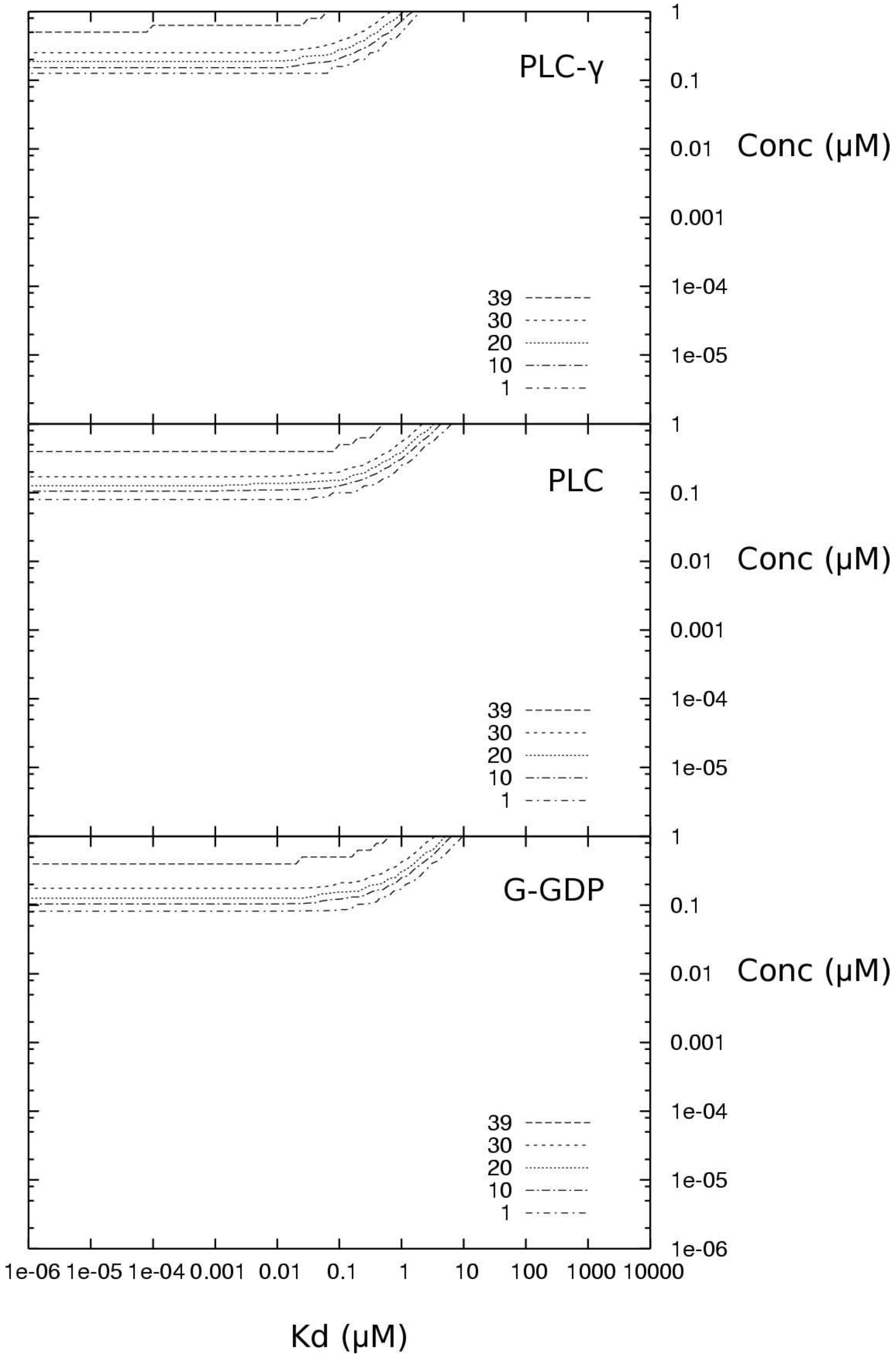


Figure 2.2: Contour plot of the response map to PLC-γ, PLC, and G-GDP as a function of drug concentration and binding affinity. In the top left portion of the graph, all mutated pathways have been successfully repressed. On the other side, none of the cell lines are inhibited. At each iso-contour line, an equal number of pathways have been successfully repressed. These targets only work at relatively high concentrations.

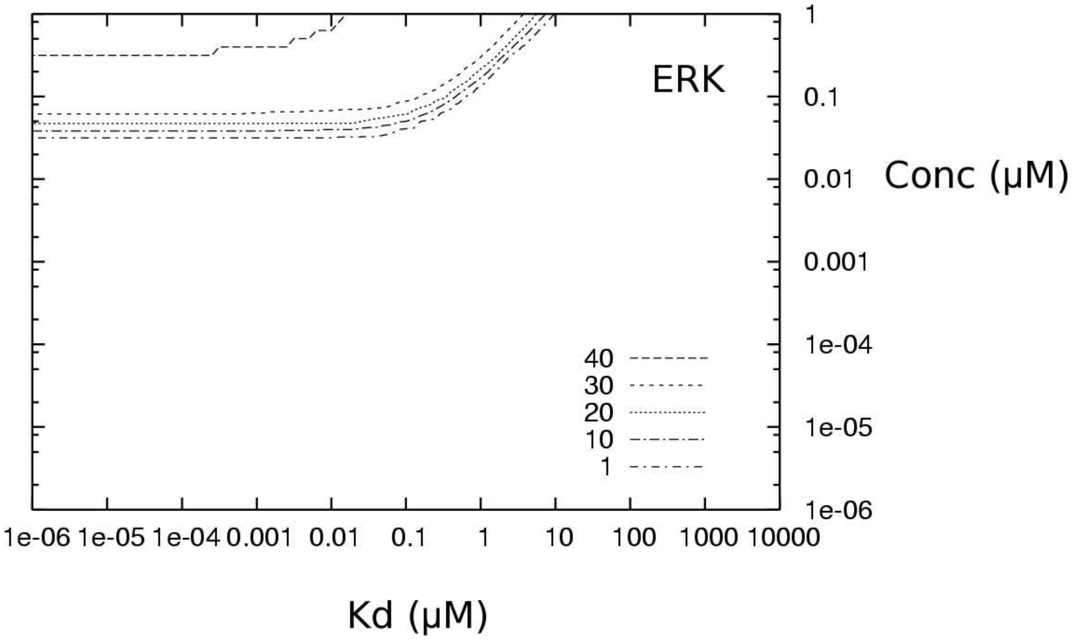


Figure 2.3: Contour plot of the response map to ERK. Drugs targeting ERK require a lower concentration and have a similar binding affinity profile as the targets in Figure 2.2.

Table 2.1: Set of deactivating mutations. The ranking of the interaction mutations is based on the highest value of k_{mod}, below which ERK activation occurs without stimulation in the MAPK pathway. Reactions with X (or X*) or (g) represent reactions involving multiple species.

Rank	k_{mod}	*Reaction*	*Dir*
1	0.75	X* + PP2A $\rightleftharpoons$ X*-PP2A	f
2	0.75	X*-PP2A $\rightarrow$ PP2A +X	f
3	0.6	Raf* +GTP-Ras $\rightleftharpoons$ Raf-GTP-Ras*	b
4	0.55	Raf* + PP2A $\rightleftharpoons$ Raf*-PP2A	f
5	0.55	GTP-Ras + GAP $\rightleftharpoons$ GTP-Ras-GAP	f
6	0.55	GTP-Ras-GAP $\rightarrow$ GAP +GDP-Ras	f
7	0.55	AA $\rightarrow$ APC	f
8	0.55	Ca-Ca$_{pump}$ $\rightarrow$ Ca$_{pump}$ +Ca$_{ext}$	f
9	0.5	X + Raf-GTP-Ras* $\rightleftharpoons$ X-Raf-GTP-Ras*	b
10	0.5	Ca + Ca$_{pump}$ $\rightleftharpoons$ Ca-Ca$_{pump}$	f
11	0.5	Raf*-PP2A $\rightarrow$ PP2A +Raf	f
12	0.5	PKC-DAG-AA $\rightleftharpoons$ PKC-DAG-AA*	b
13	0.5	PKC-DAG +AA $\rightleftharpoons$ PKC-DAG-AA	b
14	0.5	DAG $\rightarrow$ PC	f
15	0.45	Raf + PKC(g) $\rightleftharpoons$ Raf-PKC(g)	b
16	0.45	PKC$_{cytosolic}$ +Ca $\rightleftharpoons$ PKC-Ca	b
17	0.45	PKC$_{cytosolic}$ +DAG $\rightleftharpoons$ PKC-DAG	b
18	0.4	X*-MKP1 $\rightarrow$ X + MKP1	f
19	0.4	GDP-Ras + GEF(g)* $\rightleftharpoons$ GDP-Ras-GEF(g)*	b
20	0.4	MEK* + PP2A $\rightleftharpoons$ MEK*-PP2A	f

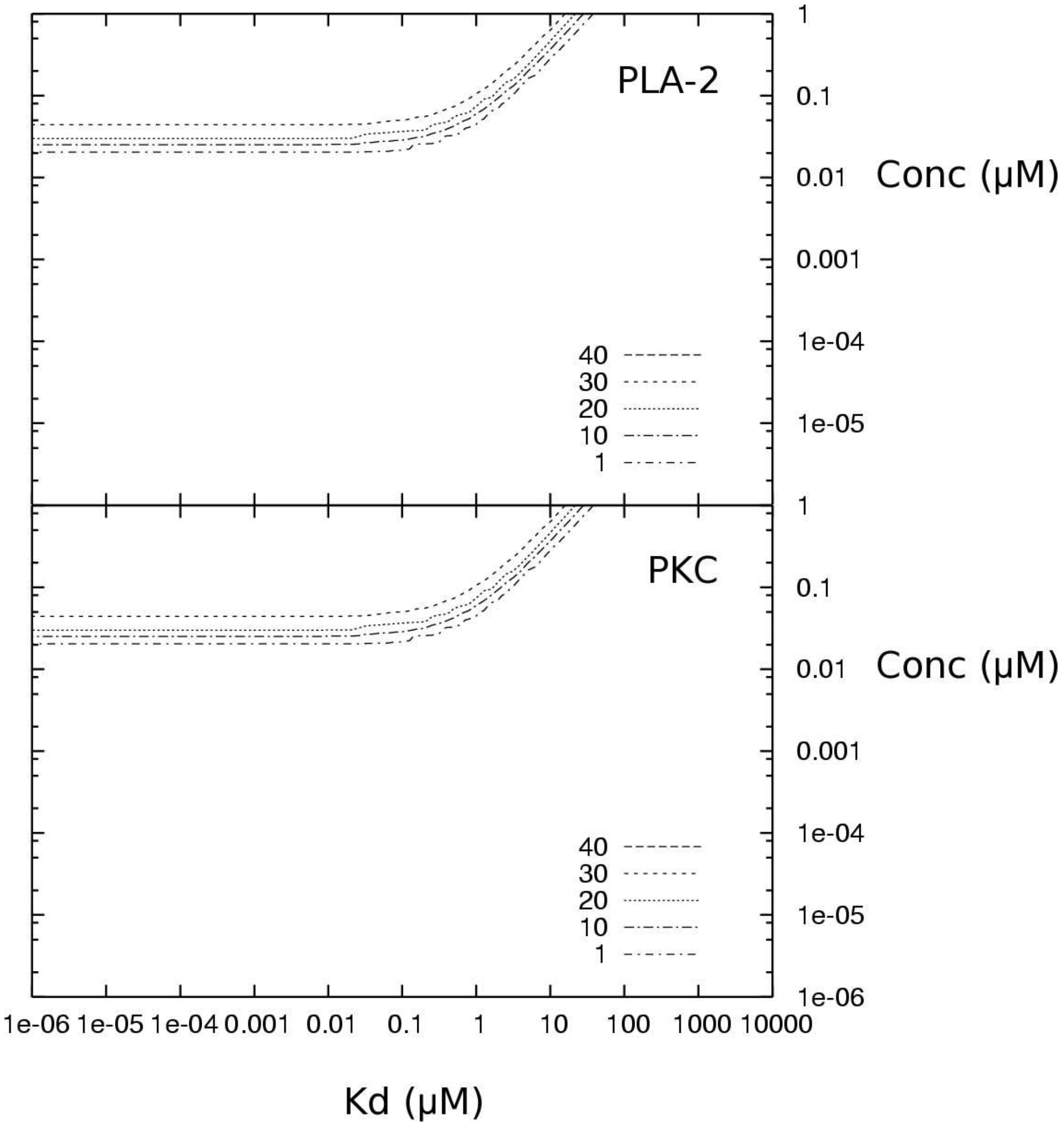


Figure 2.4: Response map of a inhibitory drug targeting PLA2 and PKC. This set of drug targets function at concentrations as low as $10^{-1.75}\mu M$.

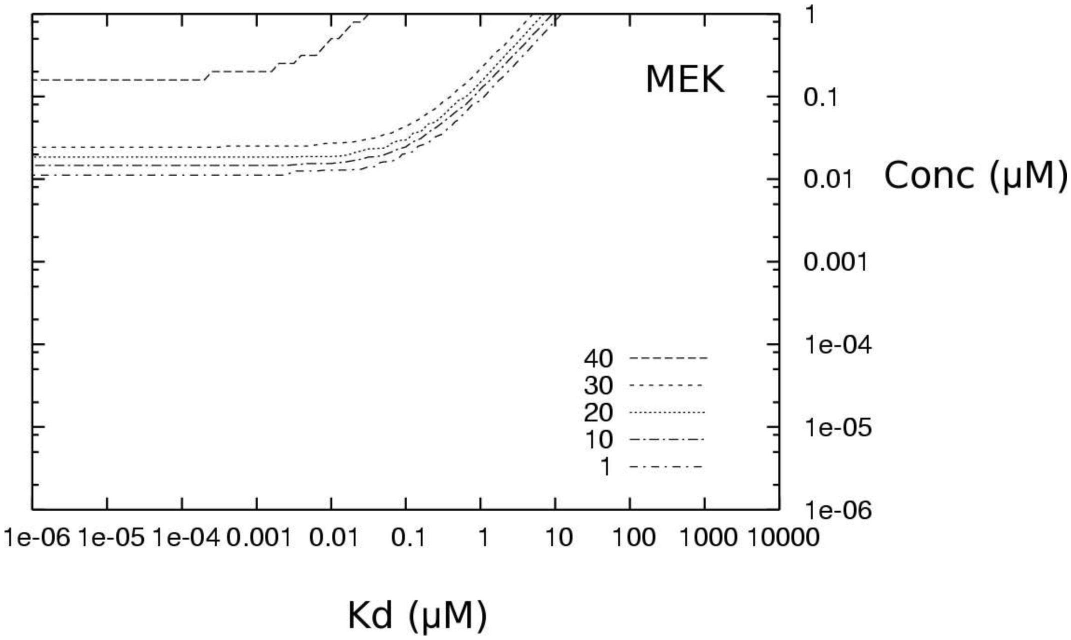


Figure 2.5: Response map of an inhibitory drug targeting MEK. While MEK can function at concentrations as low as $10^{-2.0}\mu M$, it requires higher binding affinity than targets against PKC.

Table 2.2: Set of activating mutations. The ranking of the interaction mutations is based on the lowest value of k_{mod}, above which ERK activation occurs without stimulation in the MAPK pathway. Reactions with X (or X*) or (g) represent reactions involving multiple species.

Rank	k_{mod}	*Reaction*	*Dir*
1	1.35	X + PP2A ⇌ X-PP2A	b
2	1.6	Raf-GTP-Ras* + X ⇌ Raf-GTP-Ras*-X	f
3	1.65	Raf* +GTP-Ras ⇌ Raf-GTP-Ras*	f
4	1.7	Raf-GTP-Ras*-X → X* + Raf-GTP-Ras*	f
5	1.75	GTP-Ras + GAP ⇌ GTP-Ras-GAP	b
6	1.75	Raf + PKC(g) ⇌ Raf-PKC(g)	f
7	1.85	GEF(g)* + GDP-Ras ⇌ GDP-Ras-GEF(g)*	f
8	1.9	Raf* + PP2A ⇌ Raf*-PP2A	b
9	1.9	PKC-DAG-AA ⇌ PKC-DAG-AA*	f
10	1.95	PKC-DAG +AA ⇌ PKC-DAG-AA	f
11	2	GDP-Ras-GEF(g)* → GTP-Ras + GEF(g)*	f
12	2.1	Raf-PKC(g) → Raf* + PKC(g)	f
13	2.2	Ca + Ca_{pump} ⇌ Ca-Ca_{pump}	b
14	2.2	$\text{PKC}_{cytosolic}$ +DAG ⇌ PKC-DAG	f
15	2.25	MEK* + Raf-GTP-Ras* ⇌ MEK*-Raf-GTP-Ras*	f
16	2.25	MEK** + X ⇌ X-MEK**	f
17	2.25	$\text{PKC}_{cytosolic}$ +Ca ⇌ PKC-Ca	f
18	2.3	GDP-Ras + GEF* ⇌ GDP-Ras-GEF*	f
19	2.35	GAP + PKC(g) ⇌ GAP-PKC(g)	f
20	2.4	Ca-PKC ⇌ PKC-$\text{Ca}^{*}_{membrane}$	f

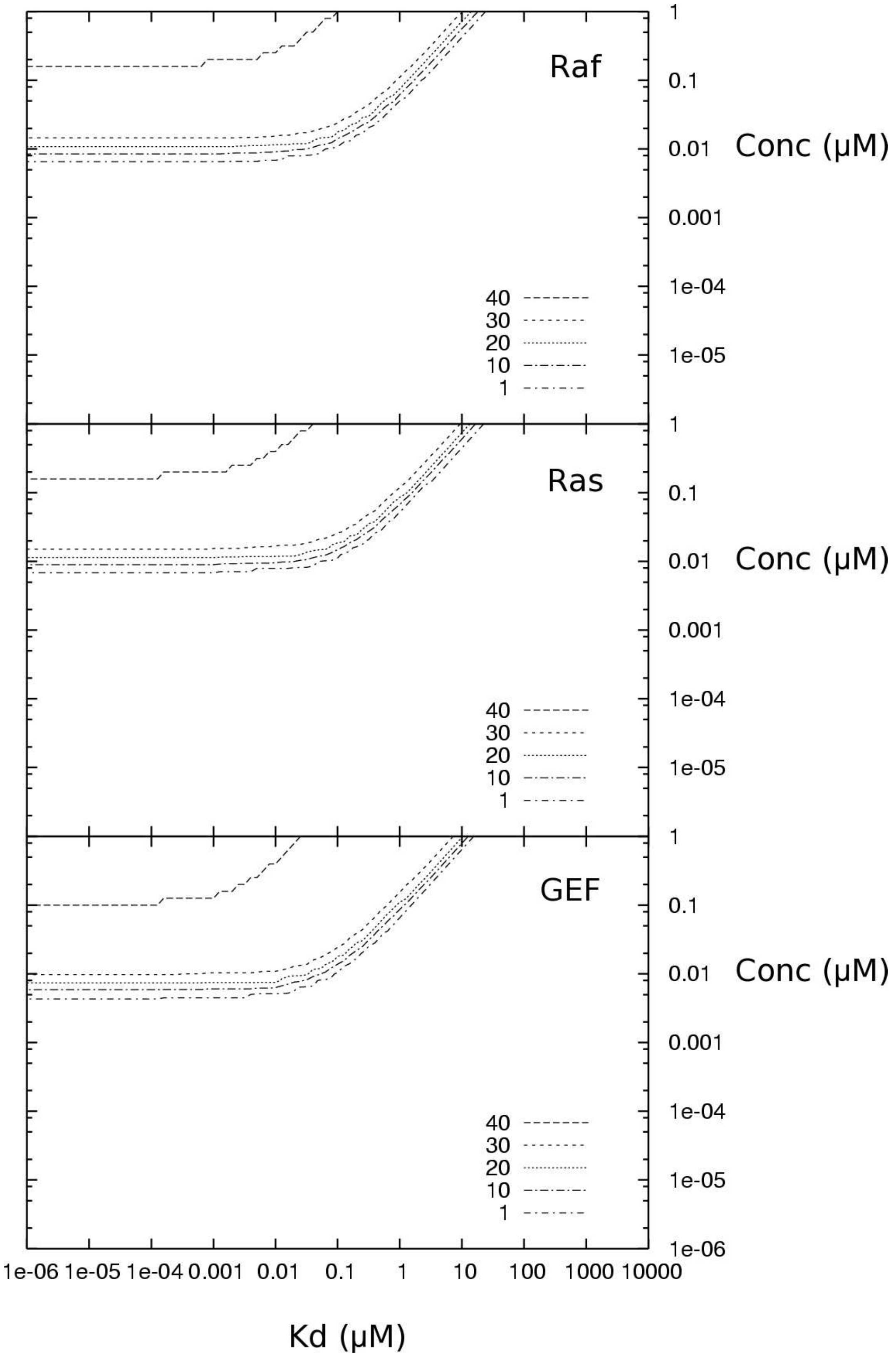


Figure 2.6: Response map of a drug targeting Raf, Ras, and GEF. This set of targets represent the best low concentration, low binding affinity drug targets found in this study.

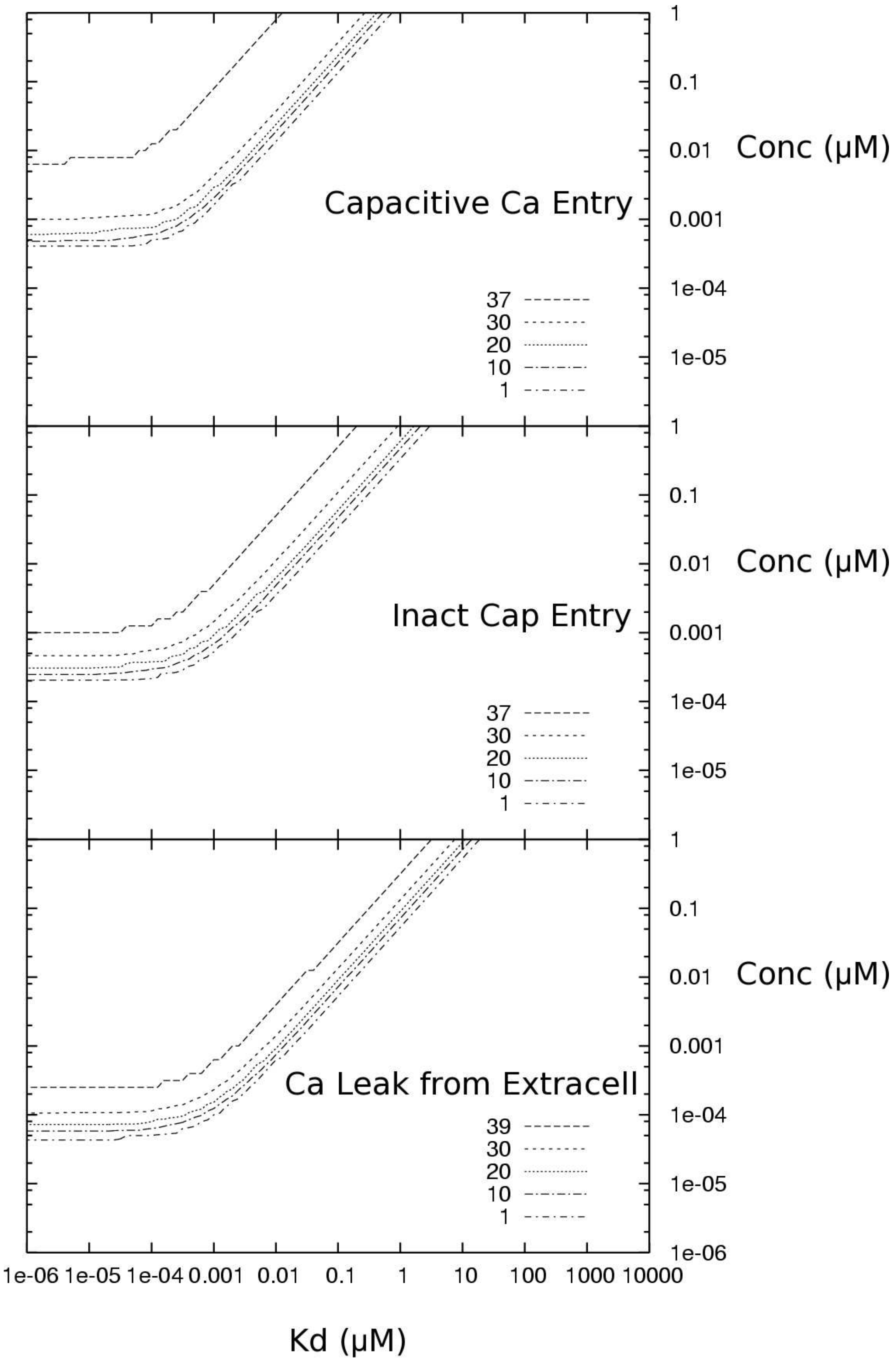


Figure 2.7: The set of calcium blockers continue to function at much lower concentrations than any other target found, however they require higher binding affinities to function properly.

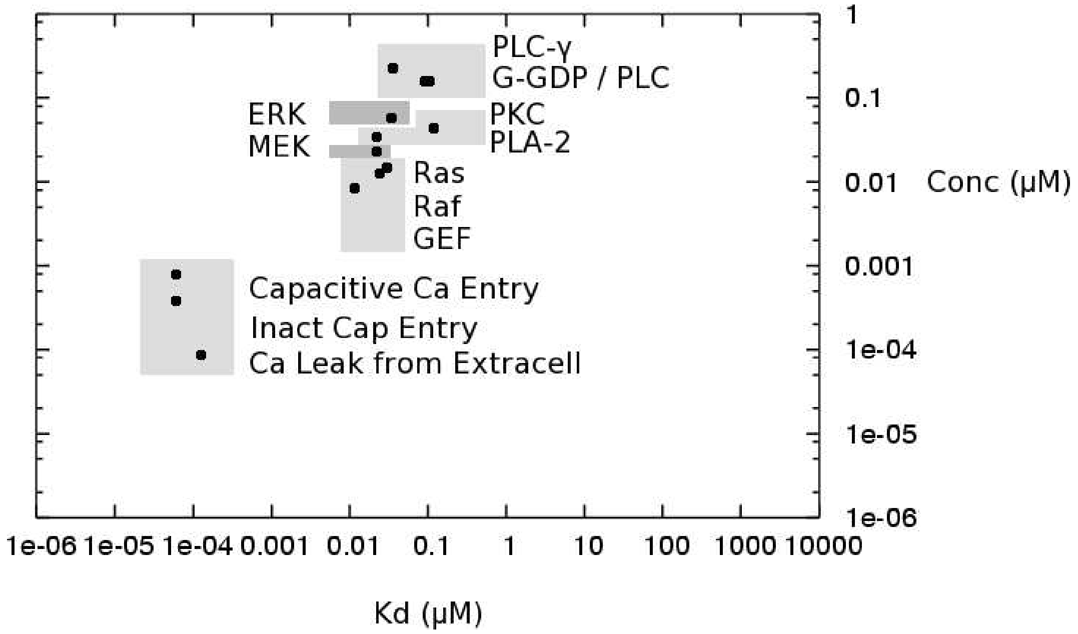


Figure 2.8: An overview of the results are shown. Each set of targets that were analyzed as a group are boxed together. Each dot represents the highest k_d at the lowest concentration that a drug (labeled by its target) needs to inhibit a majority of the mutated pathways. As there is a trade-off in binding affinity and initial concentration, a simple ranking is overly simplistic. In general, however, targets further towards the bottom right are better performing than those on the top left as drugs targeting these species function at both lower concentrations and lower binding affinities.

Chapter 3: Reducing Side Effects: A Systems Approach for the Prediction of Wild-Type MAPK Pathway Response to Targeted Drugs

3.1 Introduction

In this paper we simulate the wild-type MAPK pathway response to twelve *in silico* drugs that were previously found to reduce elevated ERK* levels in forty top ranked mutations in a model of the MAPK signal transduction pathway. We find that a subset of the simulated drugs (most notably those targeting GEF and the $G_{\alpha\beta\gamma}$ G-protein trimer) are able to effectively inhibit the elevated ERK* levels of the mutated pathways, whilst simultaneously eliciting only a small inhibition of the wild-type pathway ERK* response to EGF stimulation. While controlled MAPK pathway regulation is required for a number of wildtype cell functions (including proliferation, differentiation, and migration), excessively elevated ERK* levels have been associated with numerous cancer types, notably colo-rectal [169], breast [170], and lung [171] cancers. Moreover, MAPK pathway deregulation is found in approximately one third of all human cancers [172].

3.2 Background

In previous work [7, 6], forty top-ranked proto-oncogenes in the MAPK signal transduction pathway were identified that, when expressed, caused elevated levels of phosphorylated ERK to develop. The corresponding mutations consisted of perturbations of model parameters, specifically forward and backward kinetic rate constants. A smaller perturbation that resulted in an equal or greater increase in steady state ERK* levels was ranked higher on the list of proto-oncogenes than a larger perturbation.

Subsequently, using this list of mutations, a ranked list of drug targets within the same pathway were identified that, through competitive binding to specific pathway participants, were able to affect the signalling cascade in such a way as to reduce steadystate ERK* levels to below a chosen threshold [173, 174]. These *in silico* drugs were ranked by their minimum binding affinity and minimum initial concentration required to elicit reduced ERK* levels. A roughly ranked list of thirteen effective drugs was produced and compared to existing drugs that are either presently in clinical trials or are in some instances fully marketed cancer therapies.

As a logical extension to this work, we investigate the application of the same drugs found via the mutated pathway model to the un-mutated wild-type MAPK pathway. Will the drugs also inhibit the pathway's normally functioning ERK response to EGF stimulation or will the wild-type pathway exhibit some amount of resistance to one or more of the drugs? A therapy that is able to effectively reduce ERK* levels in mutated cells, while still allowing for normal cell function and ERK* response in non-cancerous cells would be better tolerated, and therefore more desireable, than therapies which do not distinguish between cancerous and non-cancerous cells.

3.3 Methodology

To find out if any drugs could be found that could effectively differentiate between mutated and wildtype pathways, we adopted a simulation from Figure 2d in [1], from where the original MAPK pathway model was also taken. The original figure demonstrates the pathway's ERK* response to three different EGF stimulations: a brief, high concentration stimulation; a long, low concentration stimulation; and a long, high concentration stimulation. We replicated the first of these using a 10 min, $3nM$ EGF stimulation (see Figure 3.2) using our group's `CellSim` software [8, 2].

After initially allowing the pathway to reach steady state by integrating the model equations through a one hour simulation time, a fixed $3nM$ EGF concentration is added to the model equations, sustained for 10 min, and then removed. The system is then again allowed to reach steady state. Shortly after stimulation begins, ERK* reaches a response peak of around $0.7\mu M$ concentration. Shortly thereafter, MAPK* levels fall back to basal levels via the actions of cytosolic phosphatases and EGF receptor (EGFR) internalization – a robust signal desensitization mechanism.

As described previouly, the introduction of an inhibiting drug to the system only involves the addition of a single reaction to the model's reaction set:

$$Drug + Target \rightleftarrows Drug.Target \tag{3.1}$$

The forward rate constant k_f is set to $1(\mu M \cdot s)^{-1}$ and the backward rate constant k_b (or equivalently the dissociation constant $k_d = k_b/k_f$) and initial drug concentration C_0 are taken as free parameters over which to explore. At low C_0 and high k_d values, the drug has negligible effect on the system kinetics. However at high C_0 and low k_d values, the drug in effect removes the target protein's ability to participate in the system kinetics, interrupting signaling, and results in the complete inhibition of the ERK* response.

For each of the twelve target proteins in the wild-type pathway (Figure 3.1), we ran simulations using C_0 and k_d value pairs over the region $C_0 = 10^{-6}$ to 1 μM and $k_d = 10^{-6}$ to 10^4 μM using a *log* resolution of 0.1. These points match up with those of the previous chapter. For all targets, the majority of the parameter space was of little interest and elicited either zero or total signal inhibition for both wild-type and mutated pathways. In between these two regions lay an area of transition, where the wild-type and mutated pathways exhibited responses of varying degrees of partial inhibition. As the response curves of the wild-type pathway did not identically match those of the mutated pathways, we were able to find regions of drug parameter values that elicited total or near total inhibition of the mutated pathway signaling, yet elicited little or no effect on the wild-type response.

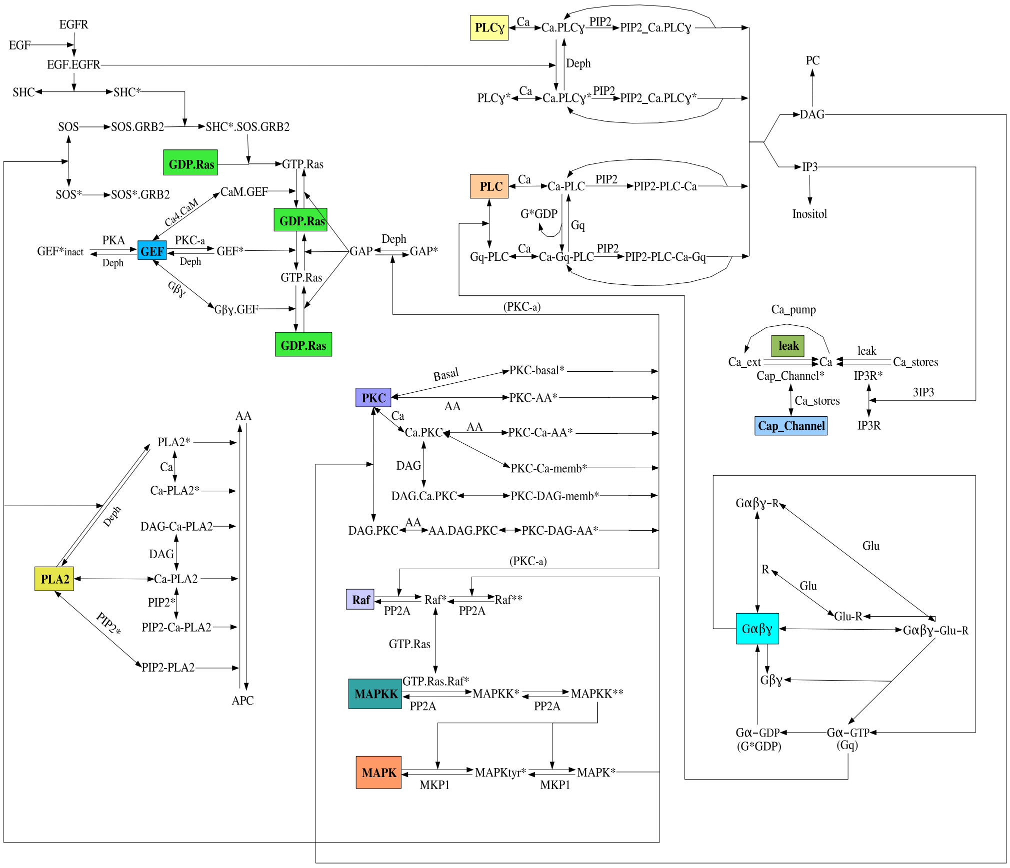


Figure 3.1: The MAPK Pathway adopted from [1] in circuit form. In this pathway, EGF binds to membrane receptors and leads to Ras activation via SHC*.SOS.GRB2 complex. The phosphorylation cascade continues via Raf and MEK, and leads to ERK activation. The twelve drug targets are highlighted in the schematic and consist of Raf, Ras, $G_{\alpha\beta\gamma}$ G-protein trimer, GEF, MEK, ERK, PLA2, PKC, PLC, PLC_{γ}, and two calcium channels.

3.4 Results and Conclusions

In dramatic fashion, Figure 3.2 illustrates the effectiveness of GEF as a drug target in a single mutated cell line. The drug reduces spurious ERK* levels in a mutated pathway (Deactivating mutation 1 in Table 2.1), while only minimally affecting the wild-type pathway's natural ERK* response to EGF stimulation.

For Figures 3.3-3.14 the peak wild type pathway ERK* response to each targeted drug is compared to each drug's effect on the forty mutated pathways. The responses are plotted over identical ranges of initial drug concentration and the drug dissociation constant $k_d = 10^{-4}\mu M$ in all cases. The plots are roughly in descending order of drug desirability, i.e. the drug's ability to deactivate the mutated pathways while leaving the wild type pathway unaffected. A mutated pathway is considered deactivated when its steady state ERK* concentration drops below $0.002\mu M$.

While these plots are useful for visualizing a drug's effect at a specific dissociation constant, they do not provide information over the full range of k_d nor do they give a quantitative measurement of the drug's ability to effectively differentiate between mutated and wild-type pathways.

Therefore, for each drug we take a sum of the normalized wild-type pathway response (ratio of peak ERK* value upon EGF stimulation to same value in an un-drugged pathway) and the normalized mutated pathway response (the fraction of the 40 mutated pathways that were successfully deactivated). This yields values between zero and two, however, as no simulated drug ever substantially inhibits ERK* in the wild-type pathway without first inhibiting the mutated pathways as well, the values do not drop significantly below 1. We therefore subtract out this lower range, yielding values in the range from zero (ERK* levels are either unaffected or completely suppressed in both mutated and wild-type pathways) to unity (complete ERK* suppression in the mutated pathways, wild-type unaffected). This yields an intuitive measure of drug desirability from undesirable (zero) to perfectly effective (unity).

We can then contour plot these desirability values as a function of initial drug concentration and dissociation constant as in Figures 3.15–3.26. The top performing targets are GEF, with a central peak region of values above 0.9, $G_{\alpha\beta\gamma}$ at 0.8, and PLC at 0.7 (Figures 3.15–3.17). The next ranked group consists of the Ca^{2+} leak channel, PKC, and PLA2, all of which reach values between 0.6 and 0.7 (Figures 3.18–3.20). The remaining six drug targets (Figures 3.21–3.26) do not elicit a value of desirability above 0.6. These targets consist of the transient receptor potential (TRP) Ca^{2+} channel, Raf, MEK, Ras, ERK, and PLC_γ.

As can be seen, the previous ranking of these targets [173, 174], based upon their ability to deactivate the mutated pathways at low drug concentrations and at high k_d values, does not identically match this alternative ranking in terms of desired wild-type response. Nonetheless, the best target in this ranking, GEF, also performed highly in the previous ranking. While all of the investigated pathway components have precedent as cancer therapy targets (as described in detail in the previous chapter), GEF's high ranking in both studies buttresses the more recent interest in GEFs as potential therapeutic targets for cancer [129].

In the pathway model used here, phosphorylated GEF acts mainly to exchange the GDP molecule bound to Ras for the more abundant GTP, activating Ras in effect [175]. However, this is not the sole role of GEFs in the cell, as GEFs constitute a large family of proteins that are involved in cytoskeletal rearrangement and motility, vesicle trafficking, and nuclear import [176]. In addition, a recent study examined the potential for small G-protein inhibition by targeting GEFs [127] and found that GEF's action of small G-protein activation is inherently spatiotemporal (i.e., localized and temporally defined within the cell). Therefore, while we have seen that purely kinetic cell signaling pathway models can describe a multitude of cell functionality and interaction, such models will need to expand in scope in order to include locality, transport, and cell geometry/morphology if they hope to accurately describe such spatiotemporal phenomena. The remaining chapters aim to describe a number of mathematical and computational tools developed by our group for use in this growing area of spatiotemporal systems biology.

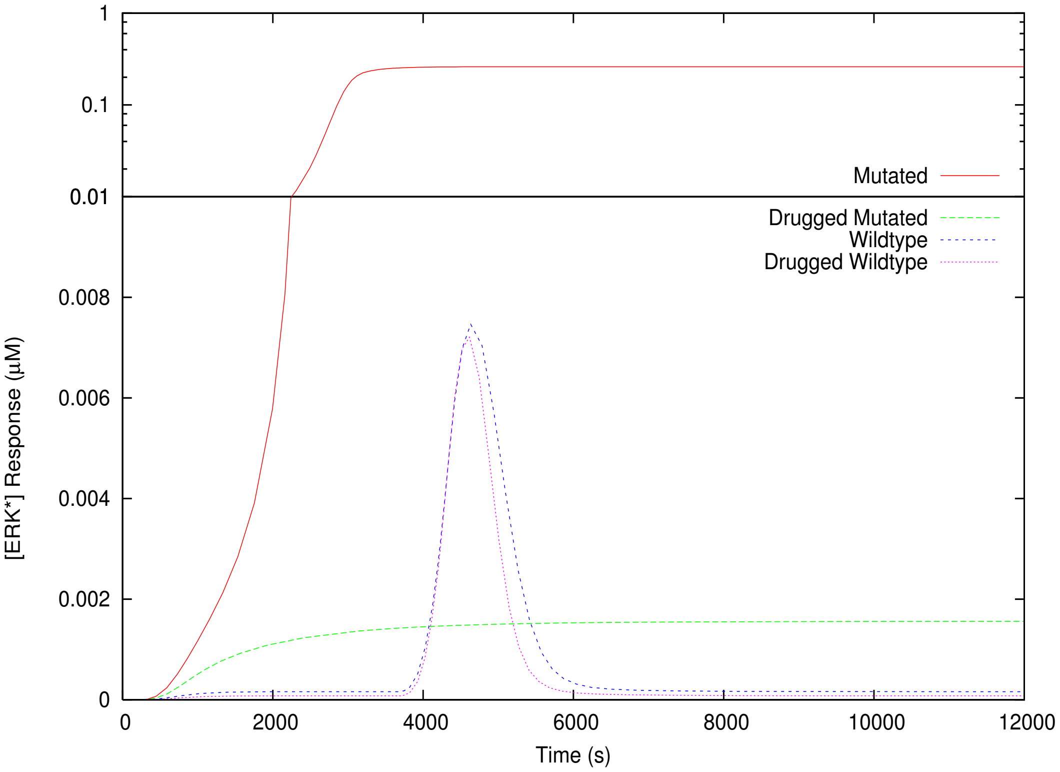


Figure 3.2: Simulated ERK* responses in wild-type vs mutated pathways. Elevated ERK* levels in a mutated pathway are reduced considerably by the addition of the drug. Conversely, ERK* response to EGF stimulation in the wild type pathway is only minimally effected by the same drug. In both cases the drug is targeting GEF with $K_d = 10^{-4}\mu M$ and $C_0 = 0.02\mu M$. EGF stimulation of the wild type pathway is 3 nM for 10 min. The mutated pathway contains mutation (1) from the set of inhibiting mutations in Table 2.1

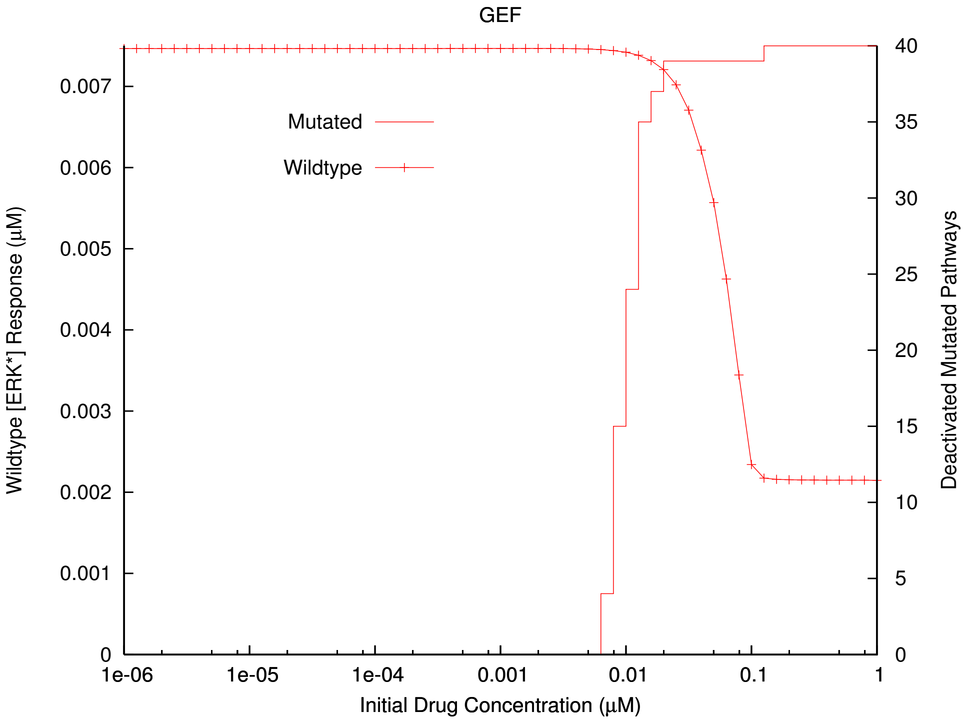


Figure 3.3: Response of both wild-type (curve, left axis) and mutated (stair step, right axis) MAPK pathways to drugs targeting GEF. The left axis corresponds to the response (peak ERK* concentration) of the drugged wild-type pathway through range of initial drug concentration. The right axis corresponds to the number of mutated pathways deactivated by the same drug through the same range of initial concentration. The drug dissociation constant is $k_d = 10^{-4} \mu M$.

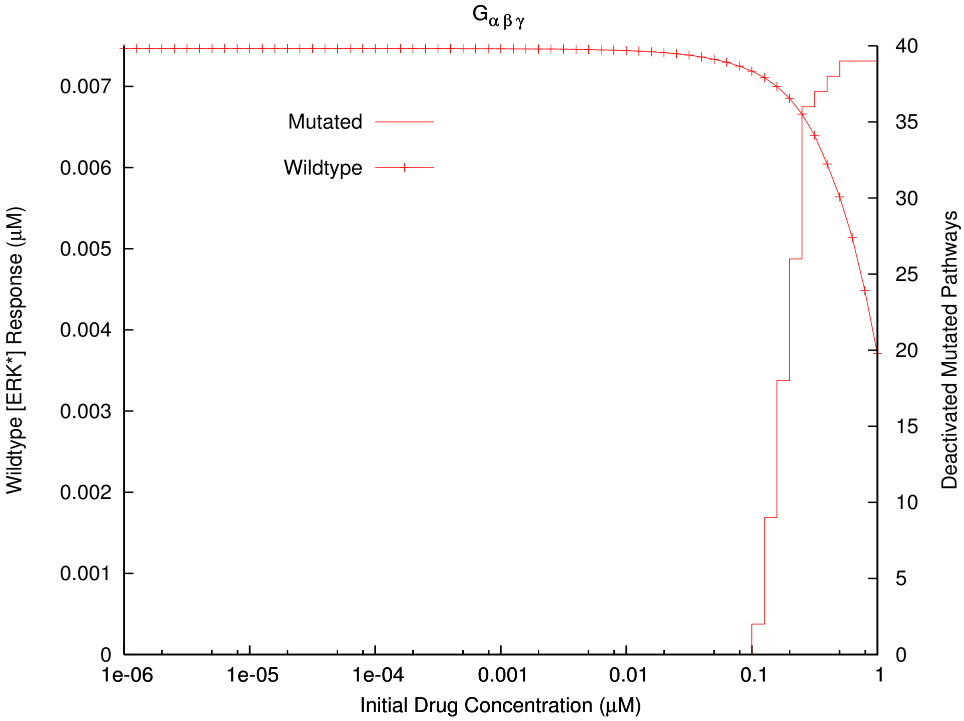


Figure 3.4: Response of both wild-type (curve, left axis) and mutated (stair step, right axis) MAPK pathways to drugs targeting $G_{\alpha\beta\gamma}$, the GDP-bound G protein trimer. The left axis corresponds to the response (peak ERK* concentration) of the drugged wild-type pathway through range of initial drug concentration. The right axis corresponds to the number of mutated pathways deactivated by the same drug through the same range of initial concentration. The drug dissociation constant is $k_d = 10^{-4} \mu M$.

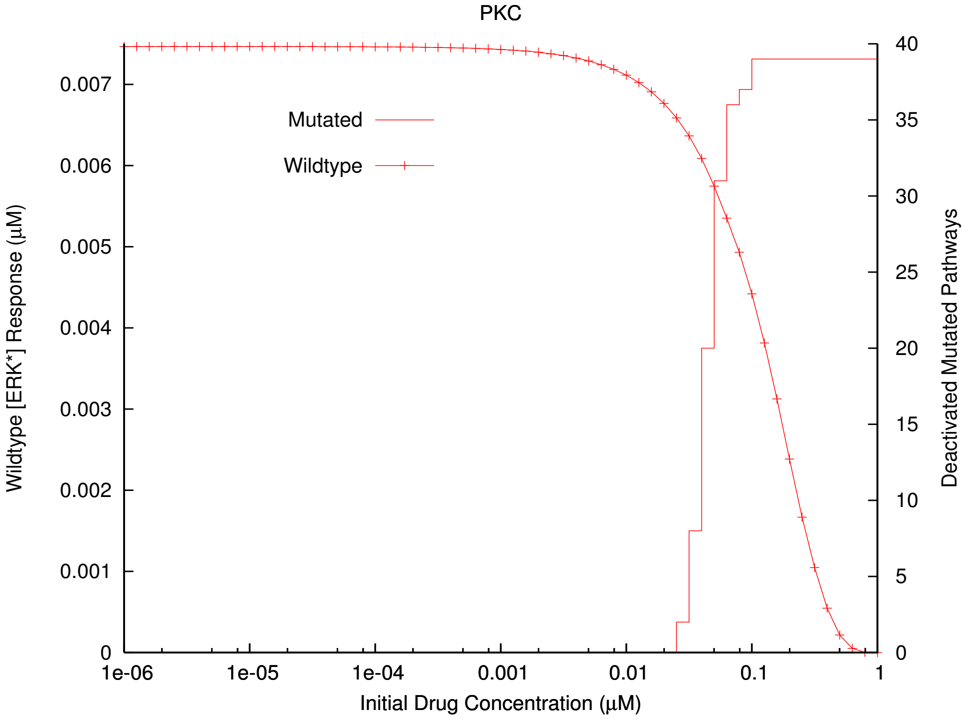


Figure 3.5: Response of both wild-type (curve, left axis) and mutated (stair step, right axis) MAPK pathways to drugs targeting PKC. The left axis corresponds to the response (peak ERK* concentration) of the drugged wild-type pathway through range of initial drug concentration. The right axis corresponds to the number of mutated pathways deactivated by the same drug through the same range of initial concentration. The drug dissociation constant is $k_d = 10^{-4} \mu M$.

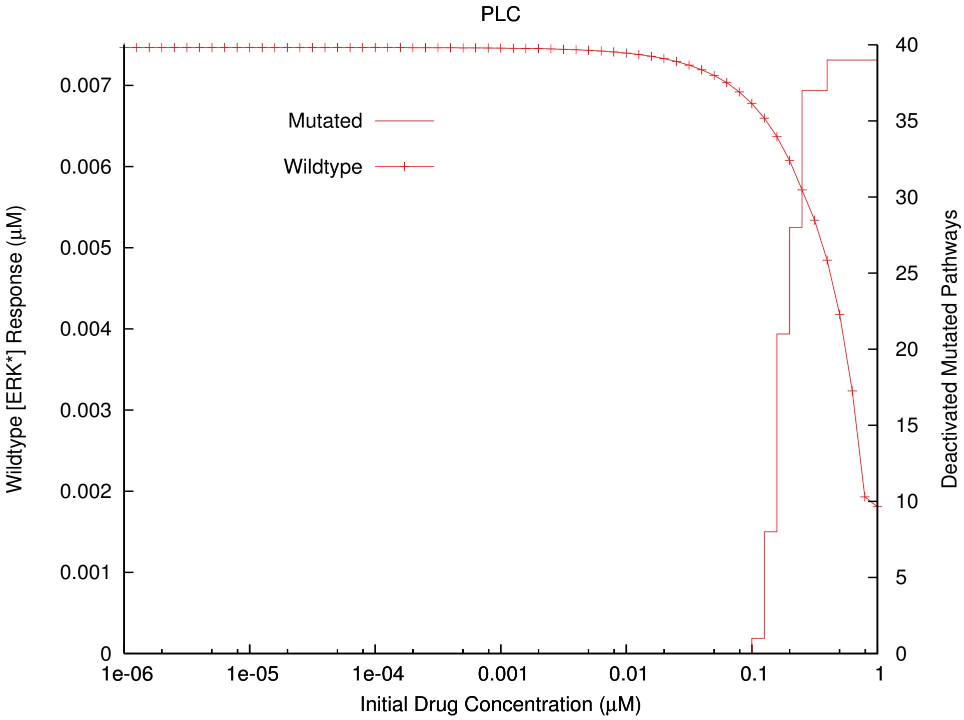


Figure 3.6: Response of both wild-type (curve, left axis) and mutated (stair step, right axis) MAPK pathways to drugs targeting PLC. The left axis corresponds to the response (peak ERK* concentration) of the drugged wild-type pathway through range of initial drug concentration. The right axis corresponds to the number of mutated pathways deactivated by the same drug through the same range of initial concentration. The drug dissociation constant is $k_d = 10^{-4}\mu M$.

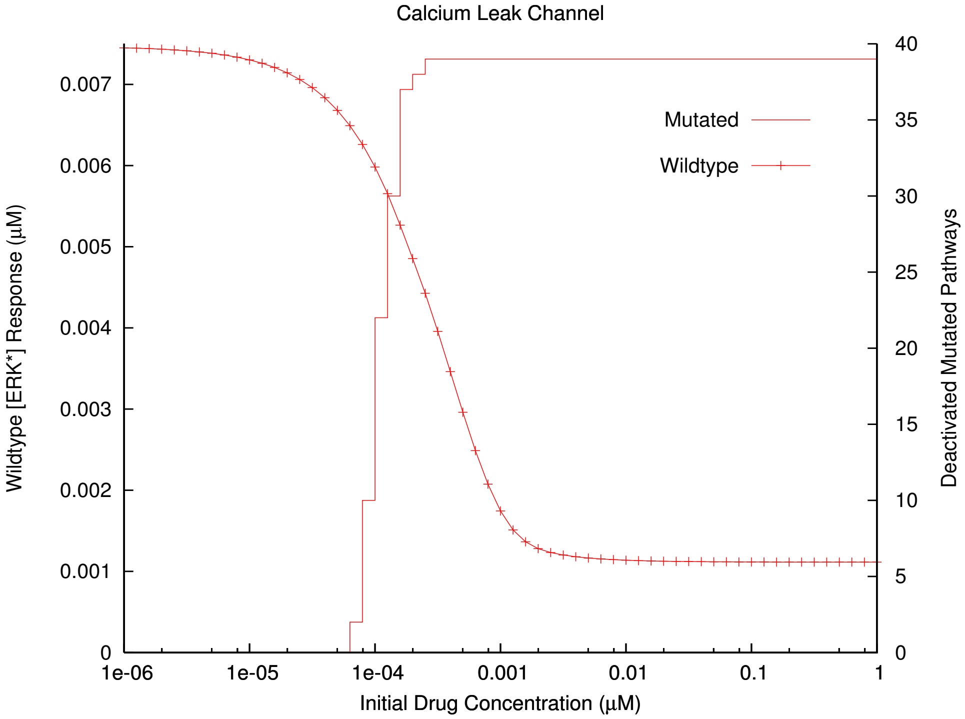


Figure 3.7: Response of both wild-type (curve, left axis) and mutated (stair step, right axis) MAPK pathways to drugs targeting the extracellular calcium leak channel. The left axis corresponds to the response (peak ERK* concentration) of the drugged wild-type pathway through range of initial drug concentration. The right axis corresponds to the number of mutated pathways deactivated by the same drug through the same range of initial concentration. The drug dissociation constant is $k_d = 10^{-4} \mu M$.

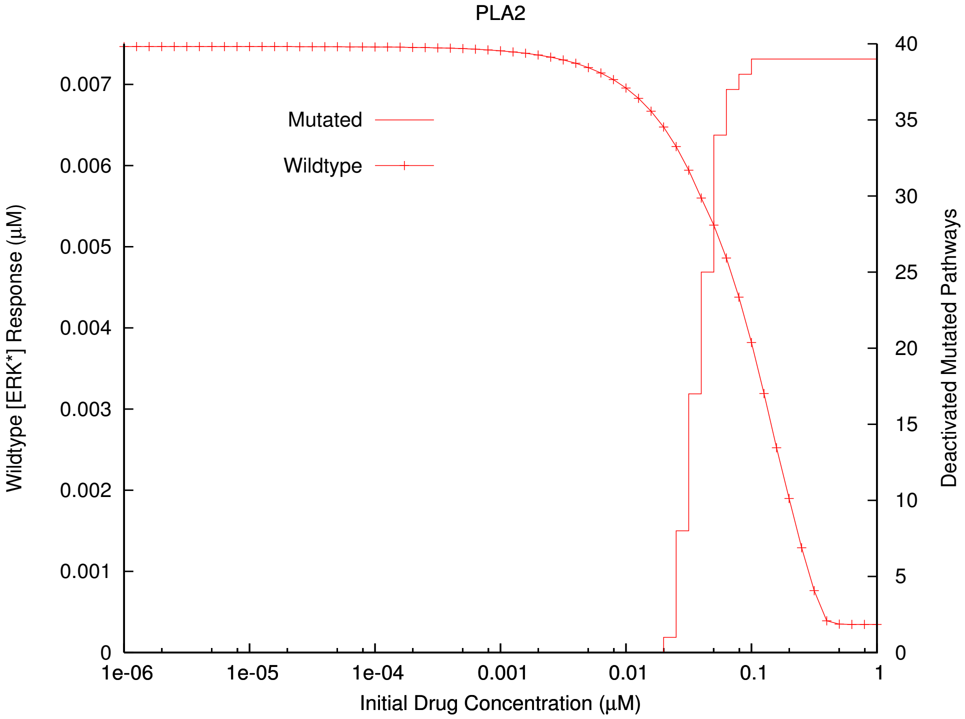


Figure 3.8: Response of both wild-type (curve, left axis) and mutated (stair step, right axis) MAPK pathways to drugs targeting PLA2. The left axis corresponds to the response (peak ERK* concentration) of the drugged wild-type pathway through range of initial drug concentration. The right axis corresponds to the number of mutated pathways deactivated by the same drug through the same range of initial concentration. The drug dissociation constant is $k_d = 10^{-4} \mu M$.

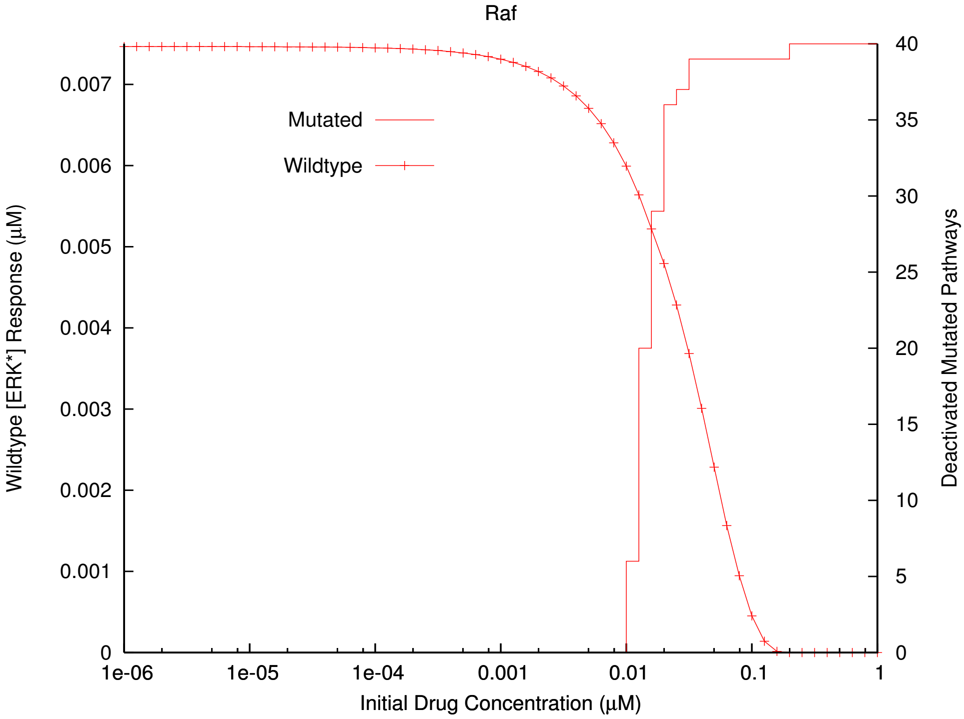


Figure 3.9: Response of both wild-type (curve, left axis) and mutated (stair step, right axis) MAPK pathways to drugs targeting Raf. The left axis corresponds to the response (peak ERK* concentration) of the drugged wild-type pathway through range of initial drug concentration. The right axis corresponds to the number of mutated pathways deactivated by the same drug through the same range of initial concentration. The drug dissociation constant is $k_d = 10^{-4} \mu M$.

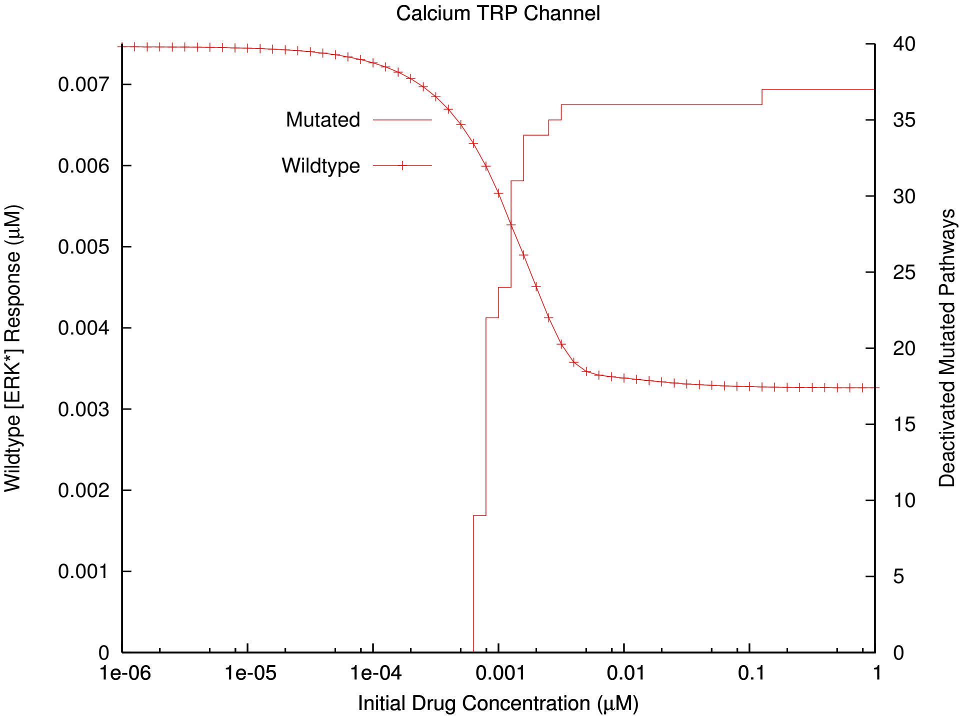


Figure 3.10: Response of both wild-type (curve, left axis) and mutated (stair step, right axis) MAPK pathways to drugs targeting the calcium TRP channel. The left axis corresponds to the response (peak ERK* concentration) of the drugged wild-type pathway through range of initial drug concentration. The right axis corresponds to the number of mutated pathways deactivated by the same drug through the same range of initial concentration. The drug dissociation constant is $k_d = 10^{-4}\mu M$.

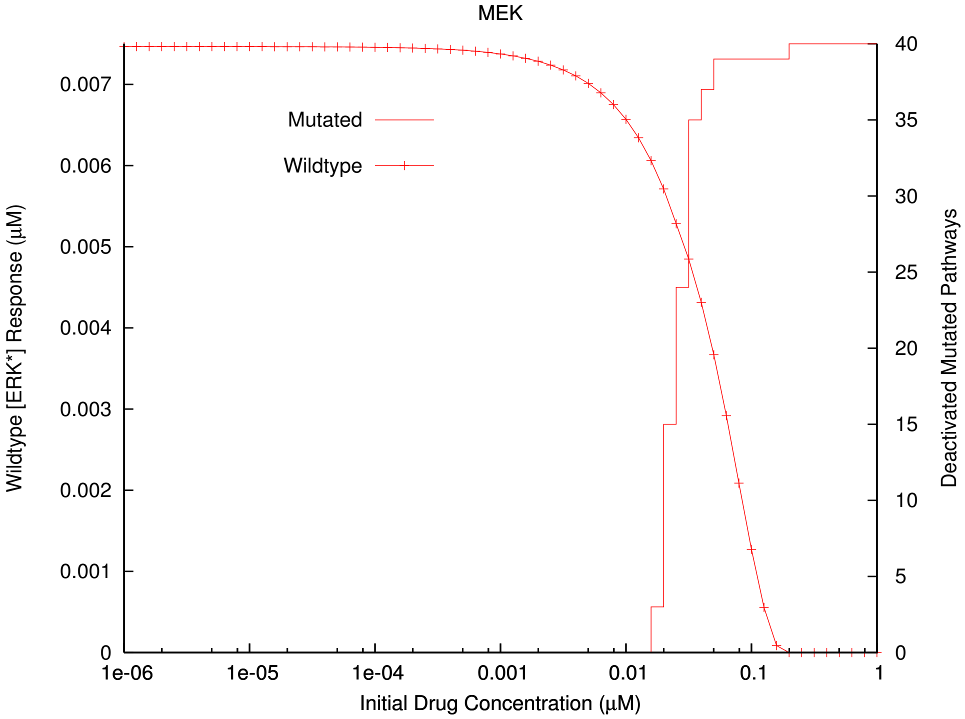


Figure 3.11: Response of both wild-type (curve, left axis) and mutated (stair step, right axis) MAPK pathways to drugs targeting MEK. The left axis corresponds to the response (peak ERK* concentration) of the drugged wild-type pathway through range of initial drug concentration. The right axis corresponds to the number of mutated pathways deactivated by the same drug through the same range of initial concentration. The drug dissociation constant is $k_d = 10^{-4}\mu M$.

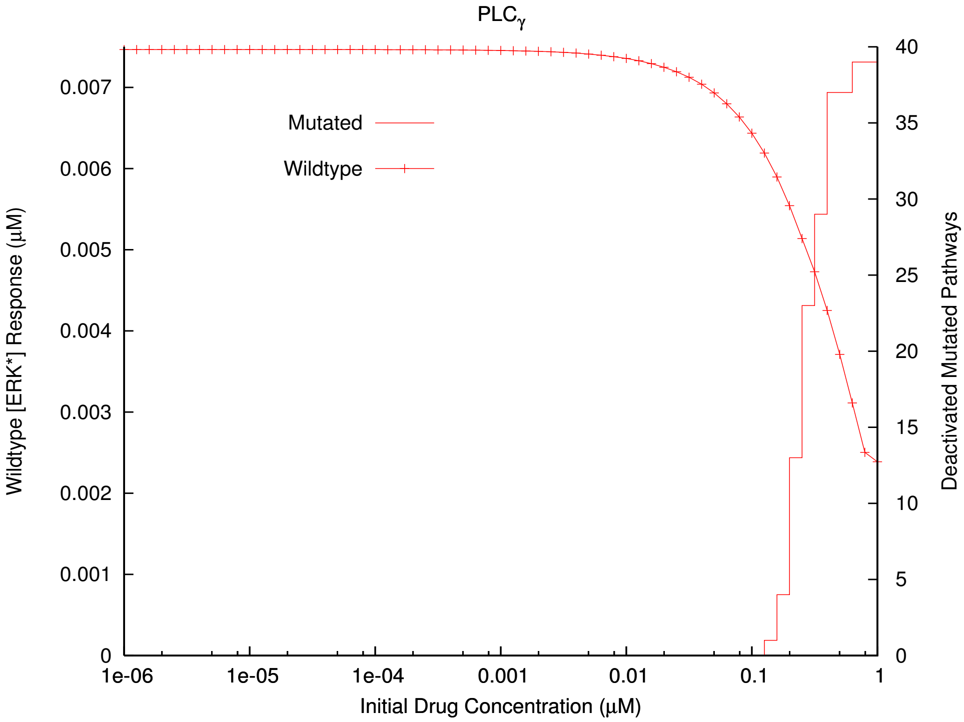


Figure 3.12: Response of both wild-type (curve, left axis) and mutated (stair step, right axis) MAPK pathways to drugs targeting PLC_γ. The left axis corresponds to the response (peak ERK* concentration) of the drugged wild-type pathway through range of initial drug concentration. The right axis corresponds to the number of mutated pathways deactivated by the same drug through the same range of initial concentration. The drug dissociation constant is $k_d = 10^{-4}\mu M$.

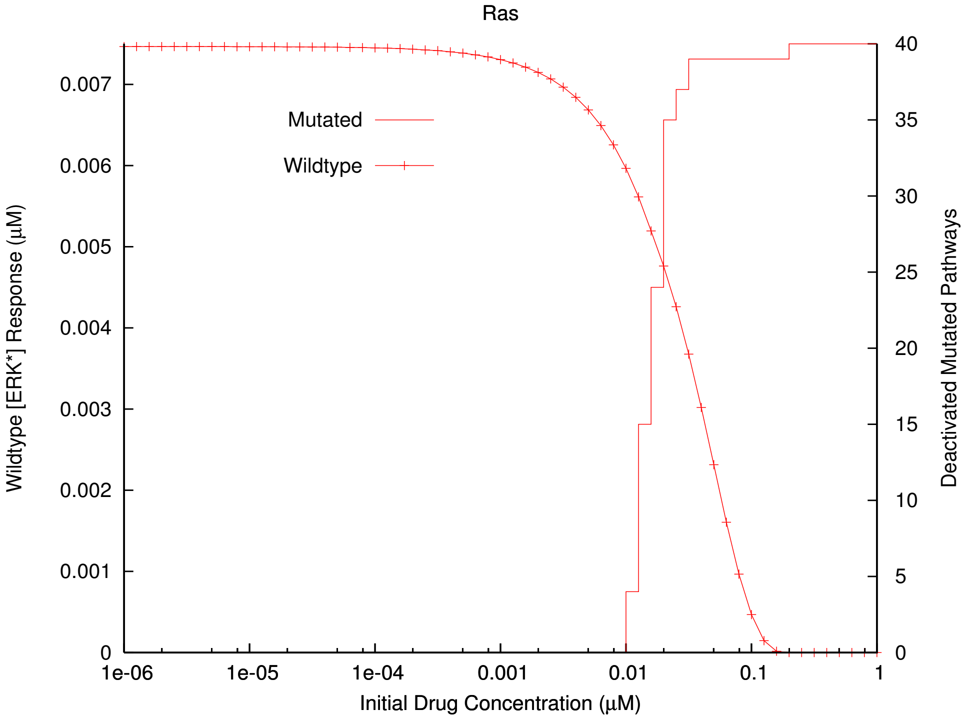


Figure 3.13: Response of both wild-type (curve, left axis) and mutated (stair step, right axis) MAPK pathways to drugs targeting RAS. The left axis corresponds to the response (peak ERK* concentration) of the drugged wild-type pathway through range of initial drug concentration. The right axis corresponds to the number of mutated pathways deactivated by the same drug through the same range of initial concentration. The drug dissociation constant is $k_d = 10^{-4}\mu M$.

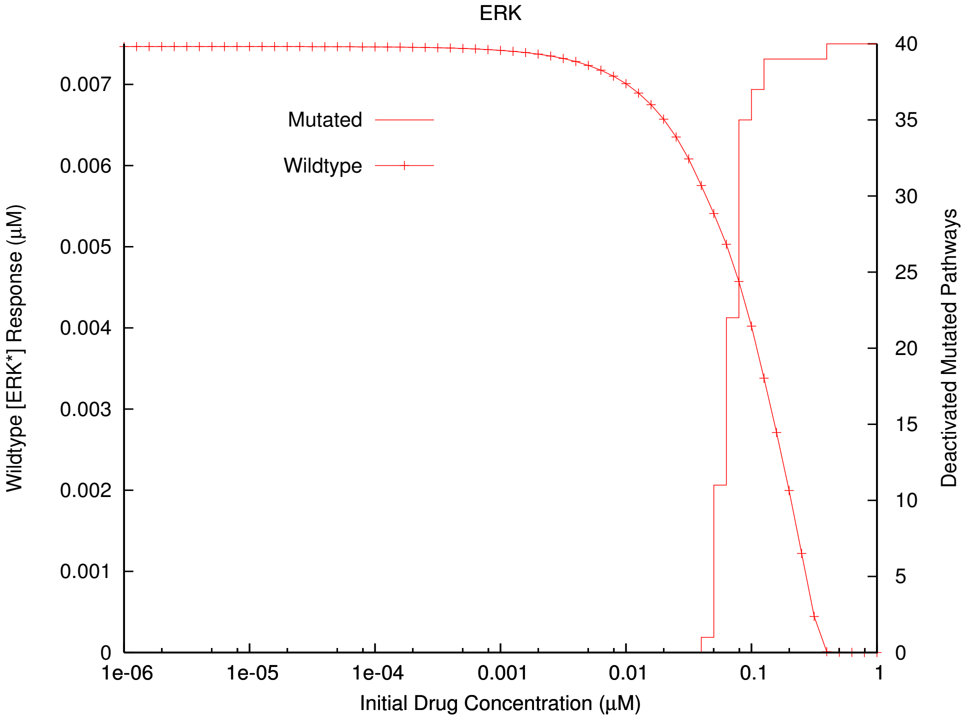


Figure 3.14: Response of both wild-type (curve, left axis) and mutated (stair step, right axis) MAPK pathways to drugs targeting ERK. The left axis corresponds to the response (peak ERK* concentration) of the drugged wild-type pathway through range of initial drug concentration. The right axis corresponds to the number of mutated pathways deactivated by the same drug through the same range of initial concentration. The drug dissociation constant is $k_d = 10^{-4}\mu M$.

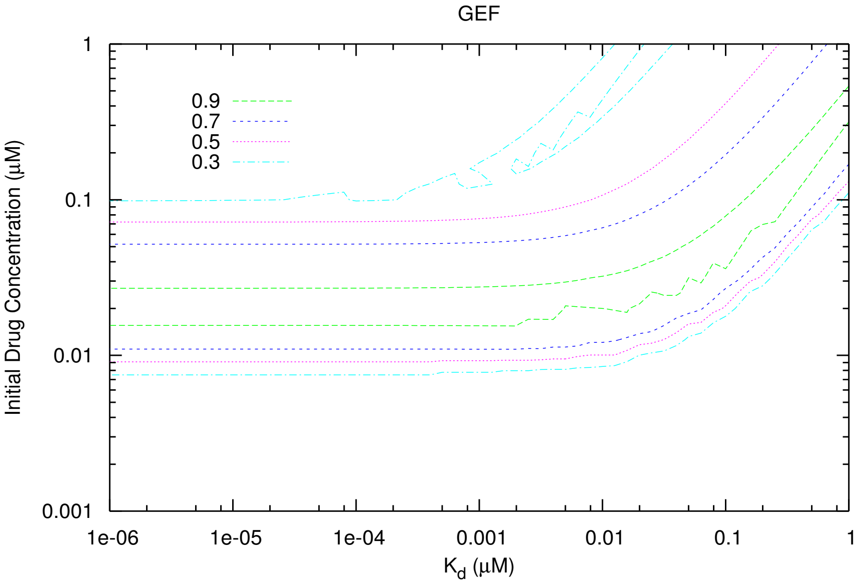


Figure 3.15: A contour plot of the sum of the normalized wild type and mutated pathway responses for the drug targeting GEF through a range of k_d and C_0. A center band encloses a region of desirability where the drug maximally affects the mutated pathways yet only minimally affects the wild type pathway. The plotted values of desirability range from zero (ERK* levels are either unaffected or completely suppressed in both mutated and wild-type pathways) to unity (ERK* suppression in the mutated pathways, wild-type unaffected).

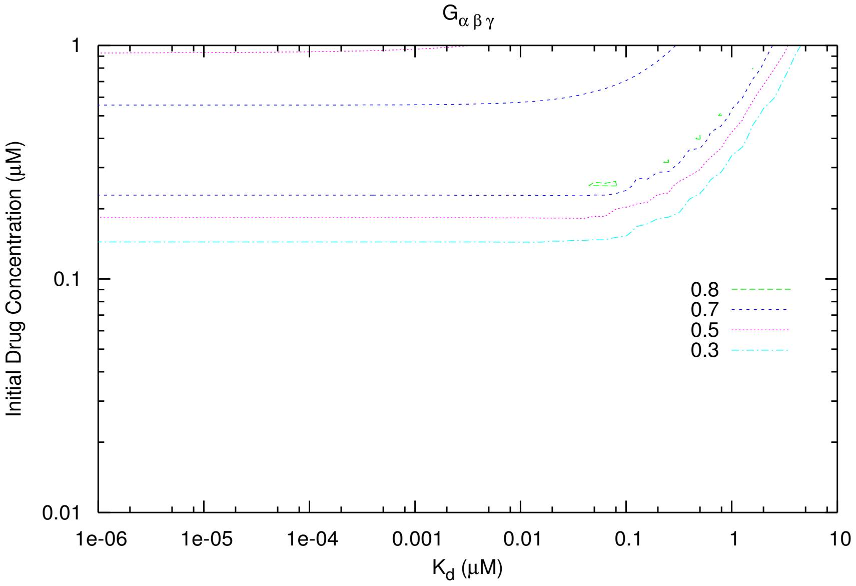


Figure 3.16: A contour plot of the sum of the normalized wild type and mutated pathway responses for the drug targeting $G_{\alpha\beta\gamma}$.

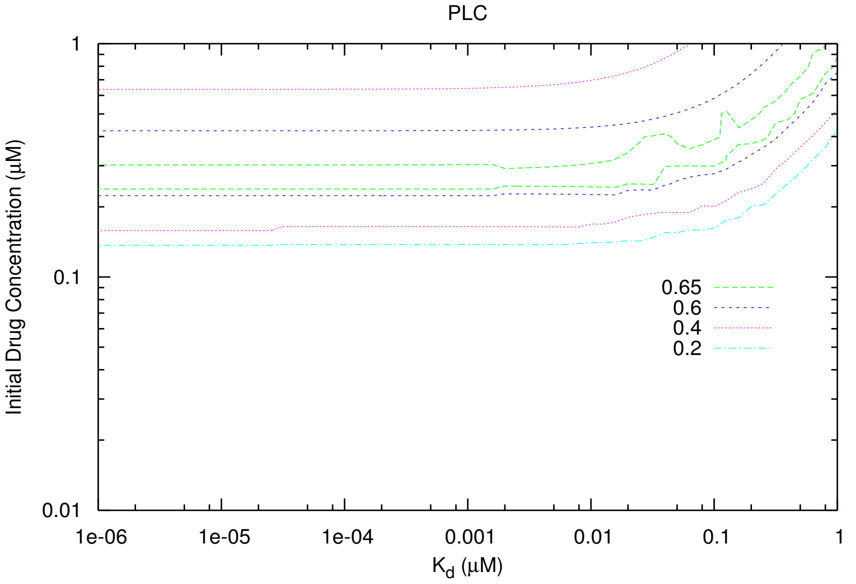


Figure 3.17: A contour plot of the sum of the normalized wild type and mutated pathway responses for the drug targeting PLC.

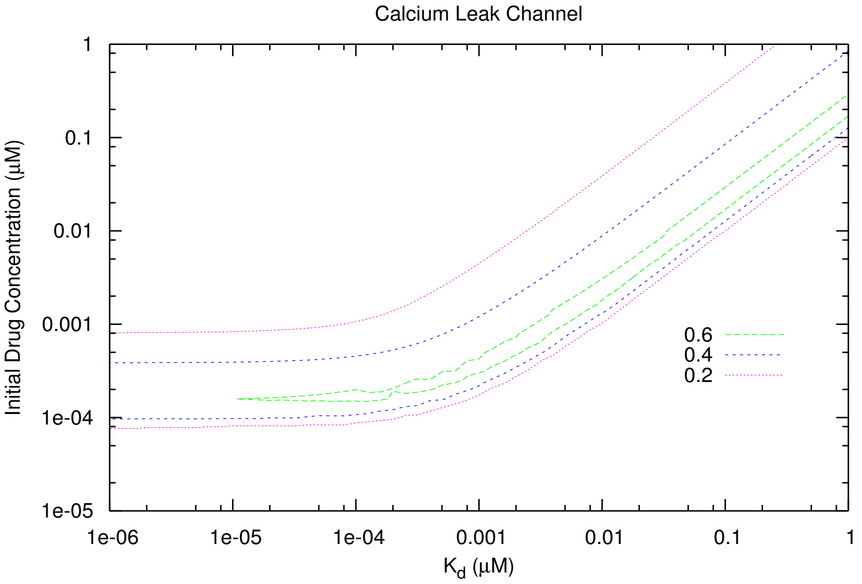


Figure 3.18: A contour plot of the sum of the normalized wild type and mutated pathway responses for the drug targeting the extracellular calcium leak channel.

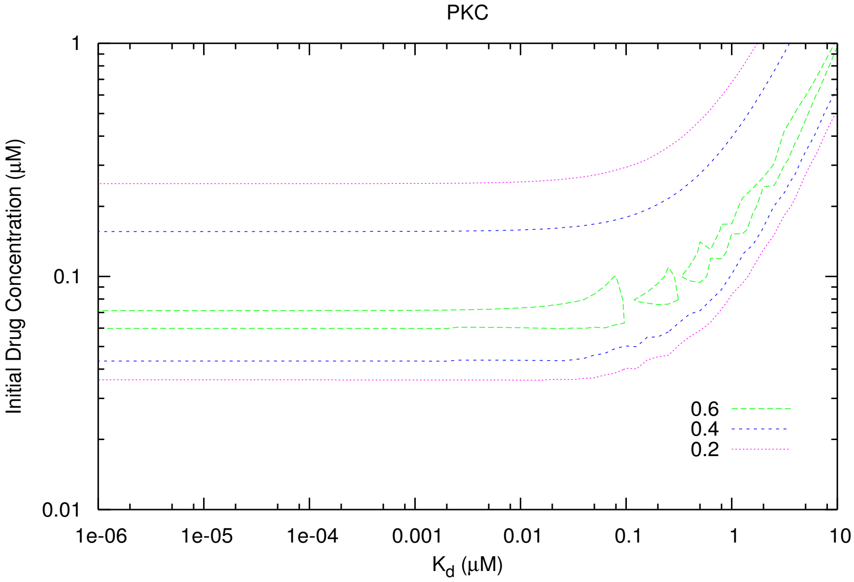


Figure 3.19: A contour plot of the sum of the normalized wild type and mutated pathway responses for the drug targeting PKC.

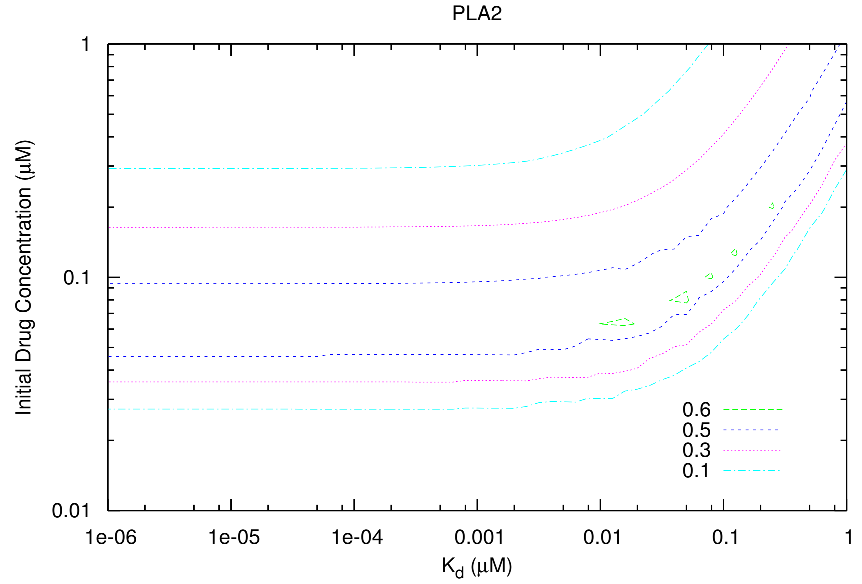


Figure 3.20: A contour plot of the sum of the normalized wild type and mutated pathway responses for the drug targeting PLA2.

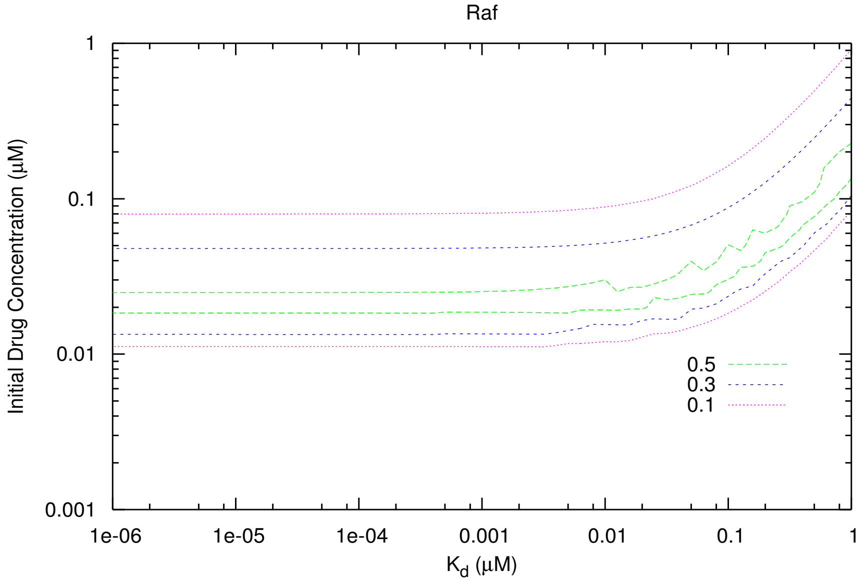


Figure 3.21: A contour plot of the sum of the normalized wild type and mutated pathway responses for the drug targeting Raf.

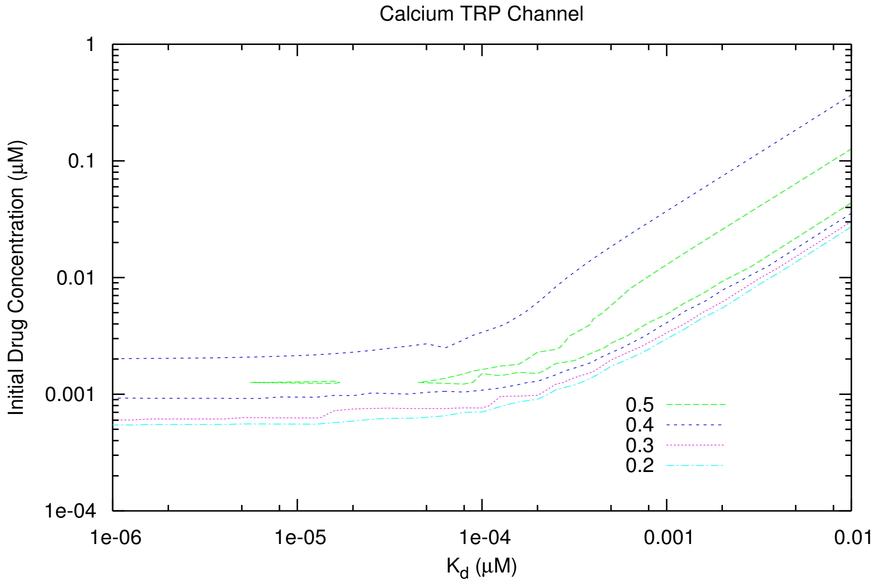


Figure 3.22: A contour plot of the sum of the normalized wild type and mutated pathway responses for the drug targeting the calcium TRP channel.

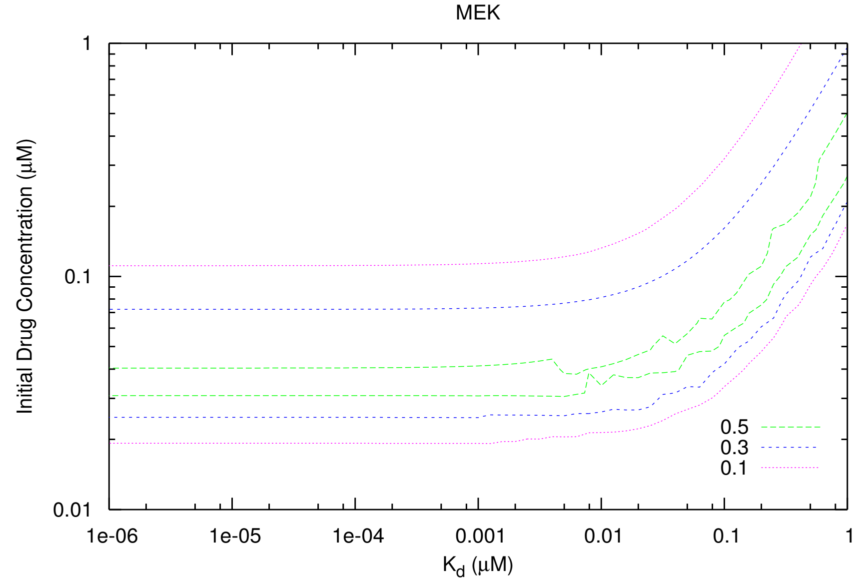


Figure 3.23: A contour plot of the sum of the normalized wild type and mutated pathway responses for the drug targeting MEK.

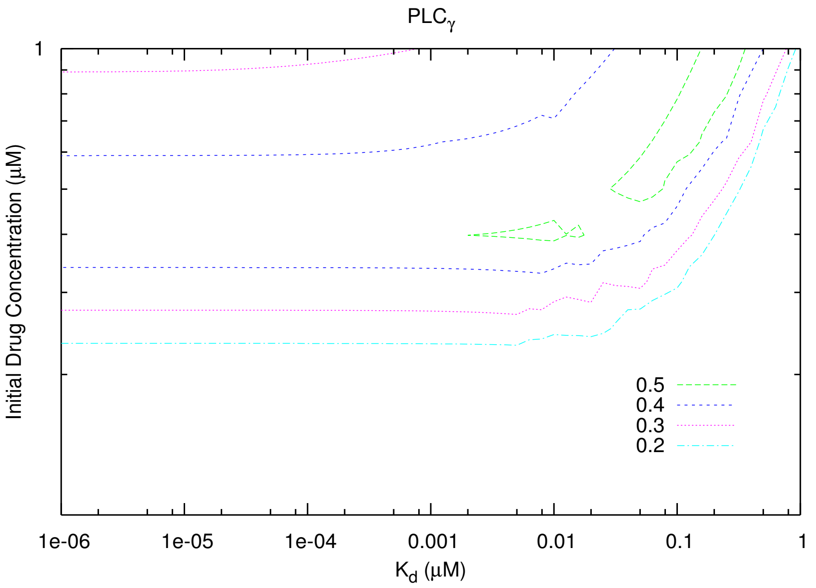


Figure 3.24: A contour plot of the sum of the normalized wild type and mutated pathway responses for the drug targeting PLC.

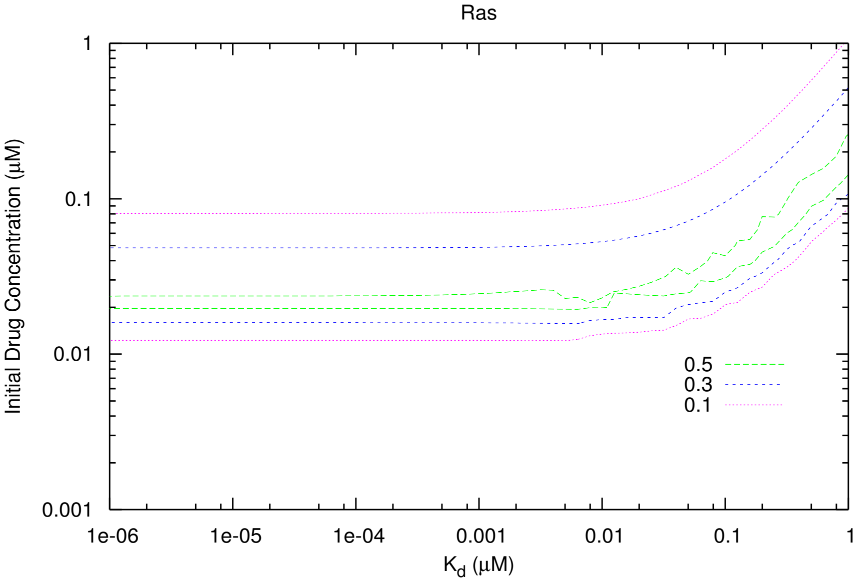


Figure 3.25: A contour plot of the sum of the normalized wild type and mutated pathway responses for the drug targeting Ras.

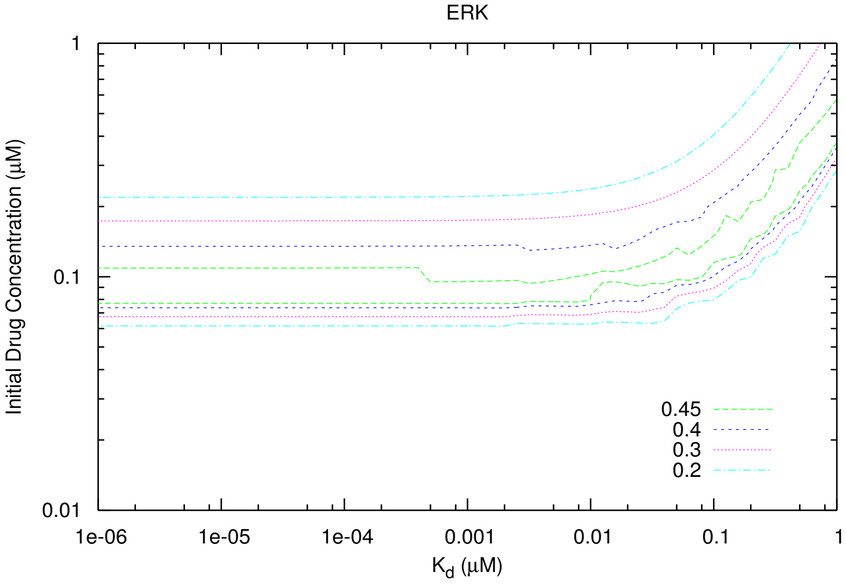


Figure 3.26: A contour plot of the sum of the normalized wild type and mutated pathway responses for the drug targeting ERK.

Chapter 4: Spatio-Temporal Systems Biology

4.1 Summary

In this chapter, we discuss computational and theoretical considerations for the extension of systems biology into the spatio-temporal realm, as well as how such an extension to systems biology may answer questions about complex protein networks and the role spatial heterogeneity may play in such processes. Both limitations and extensions of current approaches within the research community will be discussed, along with the approach taken by our group in a newly developed software package, `CellSim`. Theoretical foundations, both biophysical and numerical, are presented, afterwhich `CellSim` is introduced along with examples of simulations using the software.

4.2 Cell Compartmentalization and Heterogeneity

Systems biology has, until recently, considered the cellular activity to be fully described as simply a set of complex, coupled chemical reactions that occur concurrently to bring about the disparate and multifaceted behavior exhibited by cells. In this sense, perhaps one of the most important aspects of systems biology is its emphasis on describing the cell (and its chief component, protein) as being intimately part of this complexity, manifested in networks of protein and messenger molecule cascades. In this view, the complexity of cellular function emerges through these cascades, which may exhibit (through feedback, feed-forward, amplification and other signaling processes) important biological regulatory and functional mechanisms controlling all aspects of cellular function from metabolism to cellular growth.

It is a testament to systems biology's recent coming of age that even this immense complexity belies the true nature of cells. A cursory glimpse into the inner workings of living cells gives rise to the notion that cells are immense, heterogeneous, complex machines with a hierarchy of macroscopic ($\sim 10^{-6}m$) to microscopic ($\sim 10^{-9}m$) features acting in unison. Furthermore, cells are organized in multi-cellular systems on a much larger scale into an array of specialized and differentiated groups forming organs and other structures that encompass a viable living creature.

A host of compartments such as the mitochondria, endoplasmic reticulum (ER), nucleus, Golgi apparatus, lysosomes and peroxisomes all play important and localized roles in cellular function. The nucleus serves as a repository for the genome and is the chief location of regulatory processes controlling gene expression, as well as DNA and RNA synthesis. Synthesis of the integral membrane and secretory proteins occur within the ER and are later relocated elsewhere within the cell. The Golgi apparatus is not only a major site of carbohydrate synthesis but also provides the conduit for trafficked proteins exported from the ER. Mitochondria, which represent the energy factories of the cellular machinery, are the sources for ATP synthesis. Defunct macromolecules are degraded in lysosomes. Specific oxidative reactions that would be harmful if occurring in the cytosol are confined within peroxisomes. While the complexity of cells is inherently inscribed by the wide array of interacting protein and molecular networks and systems, the heterogeneous nature of these compartments and their interactions play a large role in regulating the protein networks thus far described. As can be seen cellular complexity is inherently spatio-temporal and more fully described as not only sets of complex protein networks within organelles and the cytosol, but as a set of interactions between compartments and the cytosol.

Protein motility within cells is guided by both passive and active transport, with protein localization controlled by specialized sorting signals (either peptides or patches). Gated transport regulates trafficking between the nucleus and the cytosol, while transmembrane protein complexes can directly transport proteins through the complex into a neighboring compartment. In addition, a large amount of soluble protein is also transported by vesicular transport. In this mechanism, a vesicle is formed in a source compartment containing the proteins to be transported and is subsequently ejected and then localized to the destination compartment. In all three cases, protein transport may be described as a combination of random motion and localized recognition via binding events. The recognition occurs through specific signal peptides or patches that may bind to a complementary recognition complex. In gated transport and transmembrane protein complexes, the complementary recognition complex is itself directly part of the transmembrane protein or the nuclear pore complex. Conversely, vesicular transport is controlled specifically by SNAREs and targeting GTPases (Rabs), which serve a similar function but will localize the entire vesicle rather than a single complex. Transport of a protein to a nuclear pore complex or to a transmembrane complex is chiefly governed by random thermal motions within the organelle itself. Similarly, localization of a vesicle to a target organelle may also be considered to be random diffusion of the vesicle coupled to SNAREs and then docked via Rabs.

4.3 Diffusion

Diffusion, the natural random motion of objects through a medium, plays a vital role in cell function in many processes such as calcium transport, transcription, and non-equilibrium dynamics [177, 178, 179, 180]. As described previously, nature has given cells numerous mechanisms for transporting materials into and out of the cell, as well as moving materials to different locations within the cell itself - notably transporter proteins, motor proteins, and transport via potential differences and ion gradients.

Typically, diffusion is neglected in most systems biology models. The model cell is instead treated as a single point in space possessing instant dilution, often called the "well-stirred" approximation. This is due to the added complexity of modeling diffusion as well as the lack of straightforward experimental techniques to provide the necessary measurements needed to fully describe a spatio-temporal model. If the time resolution of the system is large enough, this approximation is valid for many materials with fast diffusion rates and/or small volumes. Furthermore, the diffusion constant may in many cases be folded into the effective association or disassociation rate constants in Michaelis-Menten reactions. In this approximation, diffusion acts simply as a mechanism to slow down the apparent associative or disassociative rate constant, and transport between compartments may be effectively treated as exchanges between spatially averaged concentrations of the transported species. Concentration gradients of enzymes within cells that modulate signal transduction belie this simplicity [181, 182, 183, 184]. With experimental and computational technological advancements allowing finer temporal and spatial resolution, the development of spatio-temporal extensions to traditional systems biology has become much more tractable. Unless the timescale of interest is fast enough to neglect intra-compartmental concentration gradients or the concentration gradient is essentially flat, diffusion is likely to play a critical role in governing the time evolution of the system and should not be ignored.

4.4 Spatio-Temporal Systems Biology - Theory

4.4.1 The Mathematics of the Diffusion Equation

Diffusion is based on the fact that random Brownian movement [185] is statistically more likely to cause particles in areas of higher concentration to move to areas of lower concentration. One may view this as a consequence of the fact that there are simply more particles in the high concentration area that can randomly move to the low concentration area than particles in the low concentration area that can do the reverse. The mathematical equation describing diffusion is, aptly, the diffusion equation.

$$\frac{\partial \phi}{\partial t} = \kappa \nabla^2 \phi \tag{4.1}$$

This describes how the time rate of change of the amount of a substance at a location is proportional to the second spatial derivative at the same location. The expression can be derived for the case of one spatial dimension simply using elementary arguments on a Cartesian grid and can easily be expanded to higher dimensions using superposition. We shall do this here for illustrative purposes.

Assume that on a 1D grid a single particle takes a random right or left step of length dx in each time span of dt. Each step is taken to be independent of all previous steps, and the total number of particles involved is high enough to validate our probability assumptions. On average, the change in number of particles at a position x in a time step dt is given by

$$\Delta n = n_x^{t+1} - n_x^t \tag{4.2}$$

with the subscript x representing position and the superscript t representing time. For readability, n_{x-1}^{t+1} should be interpreted as the number of particles at position $x - dx$ at time $t + dt$. That is, ± 1 represents plus or minus one infinitesimal in the appropriate units. Consider the following discretization: If the particles make steps of dx each and every time step dt, and the particles have a probability p_l of moving to the left and probability p_r of moving to the right, then Δn is

$$\Delta n = p_r n_{x-1}^t - (p_l + p_r) n_x^t + p_l n_{x+1}^t \tag{4.3}$$

which simplifies to

$$\begin{aligned} \Delta n &= \frac{1}{2} n_{x-1}^t - n_x^t + \frac{1}{2} n_{x+1}^t \\ &= \frac{1}{2} \left(n_{x-1}^t - 2n_x^t + n_{x+1}^t \right) \end{aligned} \tag{4.4}$$

Multiplying through by the identity $\frac{dx^2}{dt}\frac{dt}{dx^2}$ gives

$$\frac{dn_x^t}{dt} = D \frac{n_{x-1}^t - 2n_x^t + n_{x+1}^t}{dx^2} \tag{4.5}$$

where $D = \frac{dx^2}{dt}$ equals the diffusion constant. Letting the infinitesimals go to zero while keeping D constant results in the original Equation 4.1 [186].

It is worth noting that the diffusion constant itself is dependent on a variety of factors, such as the size/shape of the diffusing particles and the viscosity/density of the diffusive medium, and must be derived experimentally [187]. For this derivation, we assume that D is not a function of position. For certain simple cases the diffusion equation may be analytically integrated. However, in general such analytic solutions do not exist for diffusion problems, and certainly not for problems pertaining to cells, where cell geometry, kinetics, and non-trivial initial conditions complicate the problem. Before embarking on this more complicated problem, we will first provide a cursory review of the coupled problem in the reaction-diffusion equation, namely the reaction portion.

The Mathematics of Chemical Kinetics

The framework of the reaction part of the reaction-diffusion equation is grounded in kinetic rate theory [188]. Every interaction between members of the signal cascade is expressed as a set of basic chemical reactions between species such as:

$$\begin{aligned} A + B &\underset{k_b}{\overset{k_f}{\rightleftharpoons}} AB & (4.6)\\ A + B &\underset{k_b}{\overset{k_f}{\rightleftharpoons}} C + D & (4.7)\end{aligned}$$

where Equation 4.6 represents an aggregation event between species A and B and Equation 4.7 represents a chemical reaction between A and B forming products C and D. The rates k_f and k_b are the forward and backward rate constants to be determined from an analysis of the response of mammalian cell assays to various perturbations. Enzymatic reactions such as phosphorylation or acetylation are represented using the Michaelis-Menten formulation:

$$E + S \underset{k_2}{\overset{k_1}{\rightleftharpoons}} E.S \xrightarrow{k_3} E + S^* \tag{4.8}$$

Such an enzymatic process is the product of two sequential processes. The catalytic step is irreversible with a rate constant of k_3 and the association is reversible with forward and backward rate constants of k_1 and k_2 respectively.

The system of kinetic reactions represented by Equations 4.6 and 4.7 can be rewritten as a series of ordinary differential equations (ODEs). These equations describe a contribution to the rate of change in concentration of a particular species as a function of time:

$$\begin{aligned} \frac{d[A]}{dt} &= k_b[AB] - k_f[A][B] & (4.9)\\ \frac{d[A]}{dt} &= k_b[C][D] - k_f[A][B] & (4.10)\end{aligned}$$

with corresponding ODEs for each of the other species expressed in Equations 4.6 and 4.7. The entire pathway is represented as a system of differential equations that describes the change in concentration of any particular species as a function of rate constants. Modeling protein interactions using only equations of the type in 4.9 and 4.10 is again referred to as the "well-stirred" approximation.

Stochastic Models

Simulating differential equations to model reaction-diffusion processes will accurately predict the average behavior of (a) large numbers of molecules within cells and (b) the average outcome of a cell process over a large number of cells. In many cases however, deterministic and continuous approaches cannot accurately simulate biological phenomena that arise from stochastic effects. For example, in the case of cancer, random molecular and cellular effects with low individual probability accumulate, eventually causing dramatic physiological effects. Biological systems, particularly those involved with genetic regulation, are very noisy, and distinct phenotypic outcomes directly result from such noise [189, 190, 191]. The problem of noise is exacerbated by the low cellular concentrations typical of many key regulatory proteins. If one speaks of nanomolar concentrations of a protein, that corresponds to just a few to tens of individual protein molecules. For example, in gene regulation there are only a few sites on DNA (each of which can be thought of as an individual "molecule" or reaction site) where transcription factors can bind and mRNA is produced. Therefore, stochastic and discrete simulations may be necessary to develop accurate reaction-diffusion models for such processes. Further discussion of such methods focusing on simulation in bacterial cells can be found in [192].

As biological processes involve a large number of molecules and protein species, the state space is too large for an exact solution of stochastic differential equations describing a reaction. Gillespie [193, 194] proposed a Monte Carlo method to exactly simulate the stochastic time evolution of a reaction system. The probability of each reaction occurring is a function of its rate constant (measured experimentally) and the number of available reactants in the simulation. At each point in time, there exists a joint probability distribution function for both the reaction and the time at which it can occur.

This generates a random trajectory through the state space that converges in the mean to the solution of the continuum model. Similarly, an average over an appropriate set of repeated experiments is expected to lead to the solution from a continuum model. In this context, one may view deterministic spatio-temporal models as the expected solutions from an appropriate ensemble average of experiments which is convenient as these ensemble averages are the simplest experimentally reproducible observables.

Arkin et al. [195] applied the Gillespie method to a fully stochastic model of E. coli infected by the λ phage virus, with two outcomes: lysogeny (integration of the phage into the bacterial DNA and “quiet” replication) and lysis (explosion of the cell and virus release). The simulation incorporated transcription and translation of genes, protein-protein and DNA-protein reactions responsible for replication, and proteases, for a total of 32 chemical reactions (including transcription and translation, which were modeled as hundreds of individual reaction events for each base). The simulation was implemented using parallel supercomputers; however, subsequent algorithmic improvements [196] have made it much faster without changing any physical assumptions.

While most applications of the Gillespie approach to stochastic reaction simulation have been only for a homogenous volume (i.e., 1D reaction systems), it has recently been applied to non-biological surface chemistry [197]. A significant drawback is scalability, since the number of time steps that must be computed increases with the total number of protein molecules to a point of intractability for eukaryotic cells. As a result, much work has been made towards developing accelerated and adaptive methods that integrate stochastic-discrete and deterministic-continuum methods at appropriate time scales. For example, the stochastically induced spatio-temporal patterns of Jung and Mayer-Kress [198, 199] has biological applications from evolution [200], electrochemical oscillators [201], neuronal models [202] and calcium signaling [203]. Turner et al. [204] is an excellent review of the state-of-the-art in stochastic biochemical simulation.

4.4.2 The Mathematics and Numerical Analysis of the Reaction-Diffusion Equation

In the spatio-temporal extension of this classical model, transport will be treated explicitly. Active transport is modeled using elementary reactions that couple to transporter proteins and may be represented by differential equations of the type 4.9 and 4.10. Passive transport can be represented with the diffusion equation for each species:

$$\frac{\partial C}{\partial t} = D\nabla^2 C \tag{4.11}$$

where D is the diffusion constant for that particular species concentration C. Active transport along actin filaments for instance may be modeled directly as part of a system of ATP driven chemical reactions.

As rate parameters need to be derived by the appropriate experimental approaches, diffusion constants may be estimated by experimental techniques such as using modulated fringe pattern photobleaching [187]. The key to building a quantitative model of the dynamical behavior of the chromatin network is coupling the system of ODEs representing the enzymatic kinetics (Equations 4.9 and 4.10) with a system of partial differential equations (Equation 4.11) representing the diffusive behavior of each species within the nucleus or on the membrane. The total contribution to the rate of change in concentration of any species at position $\vec{r}$ is the sum total of the contributions to the rate of change from all relevant reactions and transport equations. The coupling between transport and molecular kinetics may then be rewritten in a mixed finite-difference format as follows:

$$\frac{\Delta\left[C_r^i\right]}{\Delta t} = D_i\left(\frac{[C_{r+1}^i] - 2[C_r^i] + [C_{r+1}^i]}{2\Delta x}\right) + \sum_j k_{ij}[C_r^i] + \sum_{l>=m} k_{ilj}[C_r^l][C_r^m] + \sum_j \frac{P_i j}{V_i}\left([C_r^j] - [C_r^i]\right) \tag{4.12}$$

The term $[C_r^i]$ represents the concentration of species i at point $\vec{r}$ in the nuclear matrix. D_i is the diffusion constant for species i. The first sum tallies all unimolecular reactions involving species i, the second sum tallies bimolecular reactions, and the final sum represents passive diffusive transport of species i between compartments. Higher order reactions may be included in the obvious generalized fashion. In the previous equation, Δx is the spatial separation between two consecutive points, and Δt represents the temporal resolution of the numerical analysis. Using this formulation, the time evolution of each species in the protein network may be followed both spatially and temporally.

Operator Splitting

For the combined reaction-diffusion system, one may use operator splitting to propagate the total operator. Given

$$\frac{dC(t)}{dt} = L_{RD}C(t) = (L_R + L_D)C(t) \tag{4.13}$$

where L_{RD} is a reaction diffusion operator, L_R and L_D are the individual reaction and diffusion operators with corresponding propagators $U_R(t)$ and $U_D(t)$:

$$\begin{aligned} C(t + \delta t) &= U_R(\delta t)C(t) \qquad (4.14)\\ C(t + \delta t) &= U_D(\delta t)C(t) \end{aligned}$$

The 2nd order Strang splitting method [205] may be written as:

$$C(t + \delta t) = U_R\left(\frac{\delta t}{2}\right)U_D(\delta t)U_R\left(\frac{\delta t}{2}\right)C(t) \tag{4.15}$$

In a software package `CellSim`, described in more detail later in this thesis, we have implemented the Reaction-Diffusion-Reaction ordering for the splitting as recommended by Sportisse [206] and implemented by others [207]. For reaction limited models, `CellSim` uses an adaptive time step algorithm which uses a 2nd order Rosenbrock method to propagate the first operator a half step. The time step determined is then used to propagate the diffusion operator and then 2nd half of the reaction operator. It must be emphasized that this adaptive scheme is only valid for stiff reaction-limited reaction-diffusion models. A more general approach, discussed later on in this thesis, uses both the errors of each operator as well as the splitting error to determine an appropriate time step [168].

The Diffusion Operator

Many schemes exist for integration of diffusion. The most straightforward implementation of the diffusion operator is the forward time-centered space algorithm (FTCS). Using reduced units by setting the constant $a = \frac{Ddt}{dx^2}$, the FTCS method calculates the concentration n at the next time step as follows:

$$n_x^{t+1} = (1-2a)n_x^t + a(n_{x-1}^t + n_{x+1}^t) \tag{4.16}$$

Stability analysis of this algorithm reveals a stability condition of $2a < 1$. While this method is simple and stable for small time steps, it is generally inefficient and undesirable. To remove the stability condition, one could use a first order implicit scheme where the Laplacian is applied a step ahead of the current time,

$$\frac{n_x^{t+1} - n_x^t}{dt} = D\left(\frac{n_{x-1}^{t+1} - 2n_x^{t+1} + n_{x+1}^{t+1}}{dx^2}\right) \tag{4.17}$$

If spatial boundary conditions (Dirichlet or Neumann), are known then the set of equations produced by the above equation can be solved iteratively. Such solution by recursion is typical of implicit methods, where concentrations at a forward time step appear on the right side of the equation. Related to this approach is the 2nd order scheme Crank-Nicolson [208], which has a simple description as the average of the previous two methods:

$$\frac{n_x^{t+1} - n_x^t}{dt} = \frac{D}{2}\left(\frac{n_{x-1}^t - 2n_x^t + n_{x+1}^t}{dx^2} + \frac{n_{x-1}^{t+1} - 2n_x^{t+1} + n_{x+1}^{t+1}}{dx^2}\right) \tag{4.18}$$

Crank-Nicolson is unconditionally stable for dt and dx, and yields second-order accuracy in time and space. Implicit methods have the main advantage of being unconditionally stable but also require a matrix inversion. For one-dimensional problems, this method requires the diagonalization of a tridiagonal matrix at each time step. While the one-dimensional case is relatively inexpensive, two and especially three-dimensional problems require solutions of considerably more complex (although still sparse) matrices. To alleviate this unwieldy structure, further splitting of the diffusion operator into three 1D operators may be used.

This involves splitting the multi-dimensional diffusion into appropriate time intervals and applying a 1D step for each direction. In two dimensions using two steps of $\frac{\delta t}{2}$, the scheme's stability properties are maintained, but this is lost in three dimensions with three $\frac{\delta t}{3}$ time steps and the scheme becomes only conditionally stable [209]. For problems in higher dimensions, an Alternating Direction Implicit (ADI) method introduced by Douglas [210] maintains unconditional stability, is second order accurate in both space and time, and is generalizable for solving diffusion problems of arbitrary dimensionality. In 3D it may be schematically written out as

$$\begin{aligned}
\frac{w^{*}-w^{t}}{\Delta t} &= \frac{\alpha}{2}\Delta_x^2(w^{*}-w^{t})+\alpha\Delta_y^2 w^{t}+\alpha\Delta_z^2 w^{t} \\
\frac{w^{**}-w^{t}}{\Delta t} &= \frac{\alpha}{2}\Delta_x^2(w^{*}-w^{t})+\frac{\alpha}{2}\Delta_y^2(w^{**}-w^{t})+\alpha\Delta_z^2 w^{t} \\
\frac{w^{t+1}-w^{t}}{\Delta t} &= \frac{\alpha}{2}\Delta_x^2(w^{*}-w^{t})+\frac{\alpha}{2}\Delta_y^2(w^{**}-w^{t})+\frac{\alpha}{2}\Delta_z^2(w^{t+1}-w^{t}) \qquad (4.19)
\end{aligned}$$

where $\alpha = D/\Delta x^2$ and Δ^2 is the simple second order finite difference along a strip of space in the direction of the subscript:

$$\Delta_x^2 w^t = w_{x-1}^t - 2w_x^t + w_{x+1}^t \qquad (4.20)$$

Subtracting (4.19-a) from (4.19-b) and (4.19-b) from (4.19-c) reduces the scheme to three tridiagonal systems of equations, each of which can be solved efficiently using elementary linear algebra. In 1D the scheme reduces to the standard Crank-Nicolson diffusion scheme [208]

The Reaction Operator

The reaction operator may be integrated using a host of standard methods. Currently `CellSim` has the following integrators:

1. Euler
2. Exponential Euler
3. 2nd and 4th order Runga-Kutta
4. Adaptive 4th order Runga-Kutta
5. 2nd and 4th order adaptive Rosenbrock

Although Euler,

$$y\left(t+\delta t\right)=\dot{y}\left(t,y(t)\right)\delta t$$

is perhaps the simplest of numerical integrators, it is neither stable nor particularly accurate. For problems in chemistry and biology, exponential Euler takes advantage of the fact that simple kinetic interactions often give rise to exponential decay functions. That is, for kinetics one frequently encounters equations of the form:

$$\frac{dy}{dt}=A-By \tag{4.21}$$

Schematically, the exponential Euler method may be written as

$$y\left(t+\delta t\right)=y\left(t\right)e^{-B\delta t}+\frac{A}{B}\left(1-e^{-B\delta t}\right) \tag{4.22}$$

Although this scheme allows for the use of larger time steps, at low concentrations this scheme suffers from some inaccuracy that will propagate through the system and therefore should be used with caution. The workhorses of ODE solvers, Runga-Kutta methods have been implemented within `CellSim` to address this problem. The commonly used fourth-order formulation uses four strategically placed evaluations of the function's derivative within a given time step dt, and a weighted average of these derivatives is used to propagate the system a full time step:

$$\begin{aligned}
k_1 &= \dot{y}\left(t,y(t)\right)\delta t\\
k_2 &= \dot{y}\left(t+\frac{\delta t}{2},y(t)+\frac{k_1}{2}\right)\delta t\\
k_3 &= \dot{y}\left(t+\frac{\delta t}{2},y(t)+\frac{k_2}{2}\right)\delta t\\
k_4 &= \dot{y}\left(t+\delta t,y(t)+k_3\right)\delta t\\
y\left(t+\delta t\right) &= y(t)+\frac{1}{6}\left(k_1+2k_2+2k_3+k_4\right) \tag{4.23}
\end{aligned}$$

For highly coupled stiff systems of nonlinear ODEs, explicit Runga-Kutta methods become less desirable. For these systems, implicit generalizations of Runga-Kutta methods, such as Rosenbrock methods, are recommended. The Rosenbrock scheme uses a Jacobian matrix of the equations to propagate the system forward in time. In other words, the sensitivity of the solution's slope to changes in other species is considered rather than just the slope of the solution itself. One such second-order Rosenbrock method has been implemented into `CellSim` [211]:

$$\begin{aligned}
\left(\mathbf{I} - \lambda \mathbf{J} \delta t\right) k_1 &= \dot{y}(y(t)) \\
\left(\mathbf{I} - \lambda \mathbf{J} \delta t\right) k_2 &= \dot{y}\left(y(t) + \frac{k_1}{2}\delta t\right) - 2k_1 \\
y\left(t + \delta t\right) &= y(t) + \frac{3}{2}k_1 \delta t + \frac{1}{2}k_2 \delta t
\end{aligned} \tag{4.24}$$

with constant λ, identity matrix $\mathbf{I}$ and Jacobian $\mathbf{J}$.

Adaptive Algorithms and Error Analysis

As many chemical systems exist as transients that rapidly equilibrate to steady state, it is natural to seek adaptive time-step algorithms. Stiff integrators such as Rosenbrock methods have been highly successful in integrating purely kinetic systems. Inherent in such schemes is the need to calculate the Jacobian at each time step. For a single grid point with no diffusion, the Jacobian is a $N \times N$ square matrix where N is the number of reactants in the simulation. Expanding this to an extended grid of multiple points (say, n grid points in any arrangement) and including diffusion will generate a Jacobian in the form of a $n \times n$ block matrix, each block itself an $N \times N$ matrix.

This matrix is highly sparse. The second order diffusion operator would only involve nearest neighbor grid points, leaving most matrix elements empty as non-adjacent grid points do not exchange mass. Unfortunately, the size of the matrix is still cost prohibitive for performing the necessary LU decomposition required by adaptive Rosenbrock methods.

A scheme for adaptive integration of such extended spatio-temporal systems has been developed by the author and implemented in Cellsim [168]. The method extends time adaptivity to the more general case of reaction-diffusion systems by determining an error estimate for Strang's method of operator splitting at each time step and then including this information in determining future time steps. This method will be explained in detail in a later chapter of this thesis.

Spatio-temporal Sensitivity Analysis

Consider the following spatio-temporal biochemical system:

$$\frac{dC}{dt} = f(C, k, t) + D\nabla^2 C \tag{4.25}$$

where C denotes N time dependent species concentrations, the kinetics component of the system is $f(C, k, t)$ with parameters k and the corresponding diffusion component is $D\nabla^2 C$. The generalized sensitivity parameter

$$S_{ij} = \frac{dC_i}{dk_j} \qquad i = 1, ..., N \quad j = 1, ..., M \tag{4.26}$$

is then:

$$\frac{d}{dt}\frac{dC_i}{dk_j} = \frac{dS_{ij}}{dt} = \sum_{l=1}^{N} \frac{\partial f_i}{\partial k_l}\frac{\partial C_l}{\partial k_j} + \frac{\partial f_i}{\partial k_j} + \nabla^2 \frac{dC_i}{dk_j} \tag{4.27}$$

Applying operator splitting, it is clear from this equation that applying the diffusion propagator to the sensitivity parameters is sufficient to account for diffusion. The final term in the equation can be determined through simple finite differencing of the sensitivity parameters. The first two terms, however, are more complicated. Cellsim's implementation uses a Rosenbrock-based method that allows adaptive time steps to be incorporated into sensitivity calculations. In practice, two types of sensitivity parameters may be calculated within `CellSim`: sensitivity parameters with respect to k and parameters with respect to individual initial concentrations. As the latter is simpler to evaluate, we shall focus this discussion on fast evaluations of parameter-based sensitivities.

The reaction component of the previous equation may be rewritten as:

$$\sum_{l} \mathbf{J}_{il} \frac{\partial C_l}{\partial k_j} + \frac{\partial f_i}{\partial k_j} \tag{4.28}$$

where $\mathbf{J}$ is the $N \times N$ Jacobian matrix. To propagate both the sensitivity parameters S_{ij} and the original set of species C_i, consider an extended biochemical system of $(M+1)N$ equations:

$$\begin{pmatrix} \frac{dC_1}{dt} \\ \vdots \\ \frac{dC_N}{dt} \\ \frac{dS_{1,1}}{dt} \\ \vdots \\ \frac{dS_{1,N}}{dt} \\ \vdots \\ \frac{dS_{N,M}}{dt} \end{pmatrix} \tag{4.29}$$

`CellSim` will integrate the coupled system using both standard integrators as well as Rosenbrock methods. For standard integrators, the propagator is reasonably simple to define. One needs only to generate the appropriate equations and consider the extended system. The Rosenbrock method requires the generation of an extended Jacobian of the new model system. This requires the generation of the Hessian (a third-order tensor containing second order derivatives) along with several other terms in the original system.

`CellSim` automatically generates these higher order terms. The computational expense of evaluating the extended Jacobian is mitigated by its sparsity and the method's ability to use large time steps, which reduces the required number of steps. To both illustrate this procedure and describe its implementation within `CellSim`, a small sample system is introduced. Consider a simple system with five species $C_1, ..., C_5$:

$$\begin{aligned} 2C_1 + C_2 &\underset{k_2}{\overset{k_1}{\rightleftharpoons}} C3 \\ C_4 &\underset{k_4}{\overset{k_3}{\rightleftharpoons}} C5 \end{aligned} \tag{4.30}$$

`CellSim` will first automatically generate the following differential equations:

$$\begin{aligned}
f_1 &= \frac{dC_1}{dt} = -k_1[C_1]^2[C_2] + k_2[C_3] \\
f_2 &= \frac{dC_2}{dt} = -k_1[C_1]^2[C_2] + k_2[C_3] \\
f_3 &= \frac{dC_3}{dt} = k_1[C_1]^2[C_2] - k_2[C_3] \\
f_4 &= \frac{dC_4}{dt} = -k_3[C_4] + k_4[C_5] \\
f_5 &= \frac{dC_5}{dt} = k_3[C_4] - k_4[C_5] \qquad (4.31)
\end{aligned}$$

Despite the degeneracy of terms appearing in the differential equations, only unique terms are generated for use during each elementary step of the reaction propagator. In this example, the four terms are as follows:

$$\begin{aligned}
t_1 &= k_1[C_1]^2[C_2] \\
t_2 &= k_2[C_3] \\
t_3 &= k_3[C_4] \\
t_4 &= k_4[C_5] \qquad (4.32)
\end{aligned}$$

Hence, a system of differential equations may be considered simply a summation over precalculated terms. Two types of terms currently exist: one for passive transport channels (described in a later section) and one for mass action kinetics termed a `kineticTerm`. A species is indexed by two integers, one for the compartment number r (row) and one for the species c (column) in that compartment. In Cellsim, each kinetic species is stored as an instance of a simple C++ class:

```
class kineticSpeciesClass {
  public:
   int r,c;
   ...
}
```

The standardized indexing of this term (r, c) is used for fast retrieval of information about the species - perhaps most importantly the current concentration value of the species. The actual concentrations are packed into a large, contiguous memory array to minimize cache misses. Within `CellSim`, a kinetic term class has the following structure:

```
class kineticTermClass: public genericTermClass {
  public:
    svector <kineticTermClass> species;
    svector <firstderivativeClass> firstderivativesforC;
    firstderivativeforKClass firstderivativeforK
    svector <secondderivativeClass> secondderivativesforC;
    svector <secondderivativeforKClass> secondderivativeforK;
    double k;
    svector <double> jacobianMultiplier;
}
```

A `svector` may be considered simply a standard C++ STL vector that has been optimized for the purposes of `CellSim`. When using methods that require the Jacobian, the partials are all pre-generated and calculated once, minimizing the number of evaluations as well as taking advantage of the sparsity of the extended Jacobian (and other objects) that need to be built. The definitions for each term in the class are as follows:

`species`

The species in a particular term. For instance, for the term t_1 the species list contains C_1, C_1, C_2. The species C_1 is stored twice in the structure because it appears in the term as $[C_1]^2$.

`firstderivativesforC`

Stores the partial derivatives of a term with respect to all the species within the term. For the term t_1, this vector stores $\frac{\partial t_1}{\partial C_1}$ and $\frac{\partial t_1}{\partial C_2}$.

`secondderivativesforC`

Contains all non-zero Hessian derivatives for this term. For the term t_1, three derivatives are stored: $\frac{\partial^2 t_1}{\partial C_1^2}, \frac{\partial^2 t_1}{\partial C_1 \partial C_2}$, and $\frac{\partial^2 t_1}{\partial C_2^2}$

`firstderivativeforK`

Stores the partial derivative of a term with respect to its kinetic rate constant. For the term t_1, the evaluated derivative is $\frac{\partial t_1}{\partial k_1}$.

`secondderivativeforK`

Stores a vector of all the mixed terms of the form $\frac{\partial^2 t}{\partial C_i \partial k}$. For t_1 the stored terms are $\frac{\partial^2 t_1}{\partial C_1 \partial k_1}$ and $\frac{\partial^2 t_1}{\partial C_2 \partial k_1}$.

`k`

The rate constant of the kinetic term. For the term t_1, the kinetic rate constant is k_1.

`jacobianMultiplier`

A precalculated coefficient for partial derivatives with respect to C_i. Two values are stored for the term t_1.

Evaluating the original Jacobian

Before extending the system to the sensitivity parameters, we perform fast evaluation of the extended Jacobian of the fully coupled system. In our example the original Jacobian is:

$$\begin{bmatrix} -2k_1C_1C_2 & -k_1C_1C_1 & k_2 & 0 & 0 \\ -2k_1C_1C_2 & -k_1C_1C_1 & k_2 & 0 & 0 \\ 2k_1C_1C_2 & k_1C_1C_1 & -k_2 & 0 & 0 \\ 0 & 0 & 0 & -k_3 & k_4 \\ 0 & 0 & 0 & k_3 & -k_4 \end{bmatrix} \tag{4.33}$$

`CellSim` pre-generates the appropriate terms by first evaluating *firstDerivativesForC* at a given time step. The Jacobian may then be evaluated directly by taking the appropriate summation of the derivatives:

$$\frac{\partial f_i}{\partial C_j} = \sum_{k}^{L} \frac{\partial t_{ik}}{C_j} \tag{4.34}$$

where L is the number of terms for equation f_i. The Jacobian itself is stored in a special sparse matrix class that only stores the non-zero elements for the calculation. For non-Rosenbrock integrators, the sparse matrix class allows `CellSim` to use fast sparse matrix multiplies to evaluate the first term in Equation 4.28. The second part of the equation is precalculated in *firstderivativeforK*. By precalculating these terms, only derivatives requested by the user script are actually calculated. Once these terms have been evaluated, the right side of the equation may be evaluated to fully propagate the reaction portion of the sensitivity parameters.

The propagation of sensitivity parameters using the method thus described works for classes of integrators such as Euler and Runga-Kutta, however the method is not particularly suitable for stiff systems of equations. For this reason, considerable time has been spent in implementing implicit integrators such as Rosenbrock for sensitivity parameters within `CellSim`.

Calculation of the extended Jacobian

From a computational standpoint, one may consider the propagation of the extended system as simply a new system with its own corresponding Jacobian, J_c. The structure of this Jacobian has a simple block matrix form:

$$\begin{bmatrix} \mathbf{J} & | & 0 & \cdots & 0 \\ --- & + & --- & --- & --- \\ \mathbf{S}(1...N,1)_{1...N} & | & \mathbf{J} & \cdots & 0 \\ \vdots & | & & \ddots & \vdots \\ \mathbf{S}(1...N,M)_{1...N} & | & 0 & & \mathbf{J} \end{bmatrix} \tag{4.35}$$

The submatrix $S(i,j)_q$ is defined as $\frac{\partial S_{ij}}{\partial C_q}$, and the vector $S(i,1...M)_{1...N}$ is defined as $[S(i,1)_{1...N}, S(i,2)_{1...N}, ..., S(i,M)_{1...N}]^T$. As J is already calculated, the only new terms that need to be calculated are the bottom-left portion of J_C.

Numerical Evaluation of J_C

An individual term in the bottom left portion of the block matrix may be written out in the following form:

$$\frac{\partial}{\partial C_q}\frac{dS_{ij}}{dt} = \sum_l \frac{\partial^2 f_i}{\partial C_l \partial C_q}\frac{\partial C_l}{\partial k_j} + \frac{\partial^2 f_i}{\partial C_q \partial k_j} \tag{4.36}$$

The form of this equation is similar to Equation (4.28), with $J = \frac{\partial f_i}{\partial C_l}$ changed to $\frac{\partial^2 f_i}{\partial C_l \partial C_q}$ and the addition of the term $\frac{\partial^2 f_i}{\partial C_q \partial k_j}$. Hence, the procedure is the same as was used in calculating the original Jacobian J, only now the previously defined `secondderivativeClass` is also precalculated before each time-step. The $\frac{\partial^2 f_i}{\partial C_q \partial k_j}$ term also is precalculated in the same manner as before. As usual, only non-zero terms and requested terms by the user script are precalculated and used to populate the final matrix. Throughout the calculation, sparse matrix classes are used to remove unnecessary matrix multiplies through out the calculation.

4.5 `CellSim` - A Cellular Simulator

The mathematics described are all implemented in the freely available software package `CellSim` developed by our group under the Gnu Public License (GPL) [2, 212]. This package is intended for high performance distributed computing platforms that use the Message Passing Interface (MPI) parallel programming library [213].

The distributed computing platform is particularly efficient for transport-coupled kinetics. The kinetic terms are essentially communication independent as they depend only on the local concentrations of each species. Furthermore, as the computational cost of transport is much lower than the kinetic components, the system is subject to only minimal communication overhead and may often be parallelized with near linear efficiency.

4.5.1 Compartmentalization

Currently, cell geometry is explicitly defined by placing grid points on a Cartesian grid. The set of compartments defined at a grid point determines which species may exist at that point in space. The appropriate set of chemical reactions at each grid point is automatically generated from the set of all possible chemical reactions and knowledge of the compartments that contain each species. From this information, the complete set of appropriate differential equations is automatically generated over the entire grid, which is optimized for each localized grid point in both terms of storage and calculation speed. The natural boundary conditions for the system are periodic, but both Dirichlet and von Neumann boundary conditions (as well as more complicated boundary conditions) may be implemented through the appropriate use of localized chemical reactions.

4.5.2 MPI Parallelization

The explicit schemes described in the previous section allow `CellSim` to be parallelized efficiency. A large simulation grid may be split into evenly sized blocks. As the reaction operator is communication independent, the only communication cost is for diffusion across facing facets of the blocks. As the computational cost of a block will scale as the number of grid points within the block (proportional to the volume of the block) and the communication cost will go as the surface area of the block, a regime may always be found in which the communication cost of the system is small in comparison to the computational cost. Within this regime, one may naturally move to larger sized systems with linear computational cost.

4.5.3 Downloading and Compiling

A current version of the software may be downloaded via anonymous `CVS` from sodium.physics.drexel.edu. In a UNIX environment, first set the `CVSROOT` environmental variable and login as follows:

```
export CVSROOT=:pserver:anonymous@bio.physics.drexel.edu:/user/local/cvs-repository
cvs login
```

Then execute the following command to retrieve the source code:

```
cvs co cellsim-src
```

The code may be compiled using standard `make`.

In the `CellSim` source directory, we have included a script titled `setup.sh` that sets up standard compilers and optimization options. Alternatively, one may further customize a build by defining a number of environmental variables in `setup.sh` using the following machine-dependent compiler flags:

OPT

Sets any optimization flags

CXX

C++ compiler

CFLAGS

Any compiler flags

LFLAGS

Link flags

L

Linker

INCDIR

Include directory.

The `CellSim` code base is platform independent and has been compiled on MacOS, Linux and SunOS using a variety of different compilers including both the Gnu compilers and the Intel high performance compilers. It is necessary to have the freely available Gnu Scientific Library (GSL) installed on your system. If a parallel-enabled version of `CellSim` is desired, MPI must be additionally installed. The default environmental variables are set by running:

```
source setup.sh
```

An MPI enabled version may be compiled by using:

```
source setup.sh 1
```

After the environmental variables have been set, the code may be compiled by using the following commands:

```
make depend
make
```

To illustrate the use of `CellSim`, several example simulations have been included which both illustrate some of the features in the current version as well their use in spatio-temporal modeling.

4.5.4 Examples

The use of `CellSim` can be best described through the use of several biologically relevant examples that highlight some of the more salient features of the software suite. `CellSim` uses a command line based scripting interface that is executed as `cellsim file.input`, where `file.input` is the input script.

2D Gray-Scott Model of Glycolysis

The first biological example is the celebrated Gray-Scott autocatalytic model of glycolysis, originally developed by Selkov [214, 215, 216]. All input file contents (including file names) are set to be case insensitive within `CellSim`. Comments may be incorporated into any input file using C, C++ or Perl comment styles.

A typical input script appears as follows:

```
useReactions reactions.input;
useGrid grid.input;
useInitConcentrations initconc.input;

printOutput output1 {
  printinfo 1;
  printgrid 10 plot/U.10.plot U;
  printgrid 10 plot/V.10.plot V;
}

diffusionConstant all  1e-4;
diffusionConstant species U  2E-5;
diffusionConstant species V  1E-5;

printSysTime;

integrate Euler {
  dt = 1;
  dx = 0.009765625;
  runtime = 2000;
  runDiffusion;
  use output1;
}

printSysTime;
exit;
```

The main input file provides the names of all the other necessary initialization files needed for a `CellSim` run. These files are required for the initialization of the 3D spatial geometry of the simulation, the initialization of species concentrations, and the description of all the chemical reactions possible amongst the species. The main input file (in this example `file.input`) defines the simulation itself, providing definitions and instructions for printing and integration.

The first three lines of `file.input` direct `CellSim` where to find definitions of the simulation reactions (`useReactions`), grid geometry (`useGrid`), and initial concentrations of the reactants (`useInitConcentration`). These accessory files will be described in more detail in the next section.

Following the definition of the accessory files is a bracketed `printOutput` block that defines a print object named `output1`, which may be used for printing during an integration run. Multiple print objects may be defined and will only be executed with a corresponding `use` command within the integrator.

The particular `printOutput` block defined above instructs `CellSim` to print a time stamp to the screen at each step via the `printinfo` command as well as print the entire grid contents of species U and V to files every 10 steps via the `printgrid` command. After the print command, the diffusion constants are then defined for the reactants of interest via the `diffusionConstant` command. The first command uses the keyword `all` to set the default diffusion constant for all species and the second and third commands set the diffusion constant individually for species U and V.

Following the definition of the diffusion constants, `CellSim` is instructed to integrate the system equations using the `integrate` command. In this particular case, `CellSim` is using the Euler method with a fixed time step $dt = 1$, a spatial resolution of $dx = 0.009765625$, and a runtime of 2000. Within the integrate command, `runDiffusion` ensures that diffusion is enabled. In addition, `use output1` instructs `CellSim` to use the print commands defined previously in the `printOutput` command. The runtime sets the simulation time, which for this example is 2000 seconds.

The final instructions to `CellSim` are to print out the machine time (`printSysTime`) used for the simulation and then exit the simulation. The user should take care to make sure that the units are all self-consistent. The chemical reactions defined within the model are located in the file `reactions.input` defined by the `useReactions` command.

In this simulation the reactions are the celebrated Gray-Scott reactions [215]. A variant of the autocatalytic Selkov model of glycolysis, the Gray-Scott reactions are:

$$U + 2V \rightarrow 3V$$

$$V \rightarrow P \tag{4.37}$$

This simple system produces a wide variety of spatio-temporal patterns sensitive to the reaction rates and diffusion constants [216]. The reactions file for this simulation reads:

```
locationlist {
  location cytosol 1;
  default cytosol;
}

reactionlist {
  U  + V + V -> V + V + V :: 1;
  V -> P :: 0.06;

  U    -> bath :: 0.05;
  bath -> U    :: 0.05;
  V    -> bath :: 0.05;
  P    -> bath :: 0.05;
}
```

The `locationList` block defines a single compartment `cytosol` and assigns to that compartment a total volume per unit grid of 1 liter. For spatio-temporal models, leave the volume unit as 1, since the actual volume is defined by the spatial geometry. The volume definition for a compartment may be used in mixed volume kinetic models where the volume is not inherently defined by the spatial arrangement of the grid. The final command, `locationlist`, defines the default compartment for all reactants as cytosol via the `default` command. In this example, all species exist in the default compartment and thus do not need to be explicitly listed within the `locationlist` command.

The next block, `reactionlist`, defines the reactions in the Gray-Scott model. All the reactions are nonreversable with rate constants following double colons. For this simulation, the grid is simply a square plane of cytosol. The grid file defined by the `useGrid` command takes the following form:

```
grid 48 48 1;
0 0 0 cytosol;
0 1 0 cytosol;
```

```
0 2 0 cytosol;
...
0 47 0 cytosol;
...
47 0 0 cytosol;
...
47 47 0 cytosol;
```

Here the grid has dimensions of 48 × 48 × 1, and all species are defined to exist in the compartment `cytosol`. A grid point may be defined to have any number of compartments. Grid points with multiple compartments may be considered to be an interface region, and reactions involving species of different compartments may additionally react within such interface regions.

The initial conditions in this example consist of a grid containing two areas, a central square and the area surrounding it. The two species U and V initially exist in both areas at different, randomly perturbed concentrations. For `CellSim`, the initial concentrations can be specified for a species throughout all its compartments or individually specified at each grid point. Using perturbed concentrations, the `initconc`.input file contains the following commands:

```
P = 0.0;
fixed bath = 1.0;
point 0 0 0  U = 1.00312899386658;
point 0 0 0  V = 0;
point 0 1 0  U = 1.00259570383182;
point 0 1 0  V = 0;
point 0 2 0  U = 0.996343013072222;
point 0 2 0  V = 0;
point 0 3 0  U = 0.990340706493643;
...
```

The first command defines the concentration of P to be 0 everywhere. The next command sets the bath to a concentration of 1.0 units and fixes it as a constant value. Species U and V are perturbed around 1.0 and 0.0, respectively, in the outer region and 0.5 and 0.25 in the inner region. All concentrations are listed individually at each grid point. As with all `CellSim` files, the scripting language will override previous commands with any subsequent commands, so a default concentration may be set and then altered at specific grid points with the `point` command.

3D Kinase Phosphatase Model

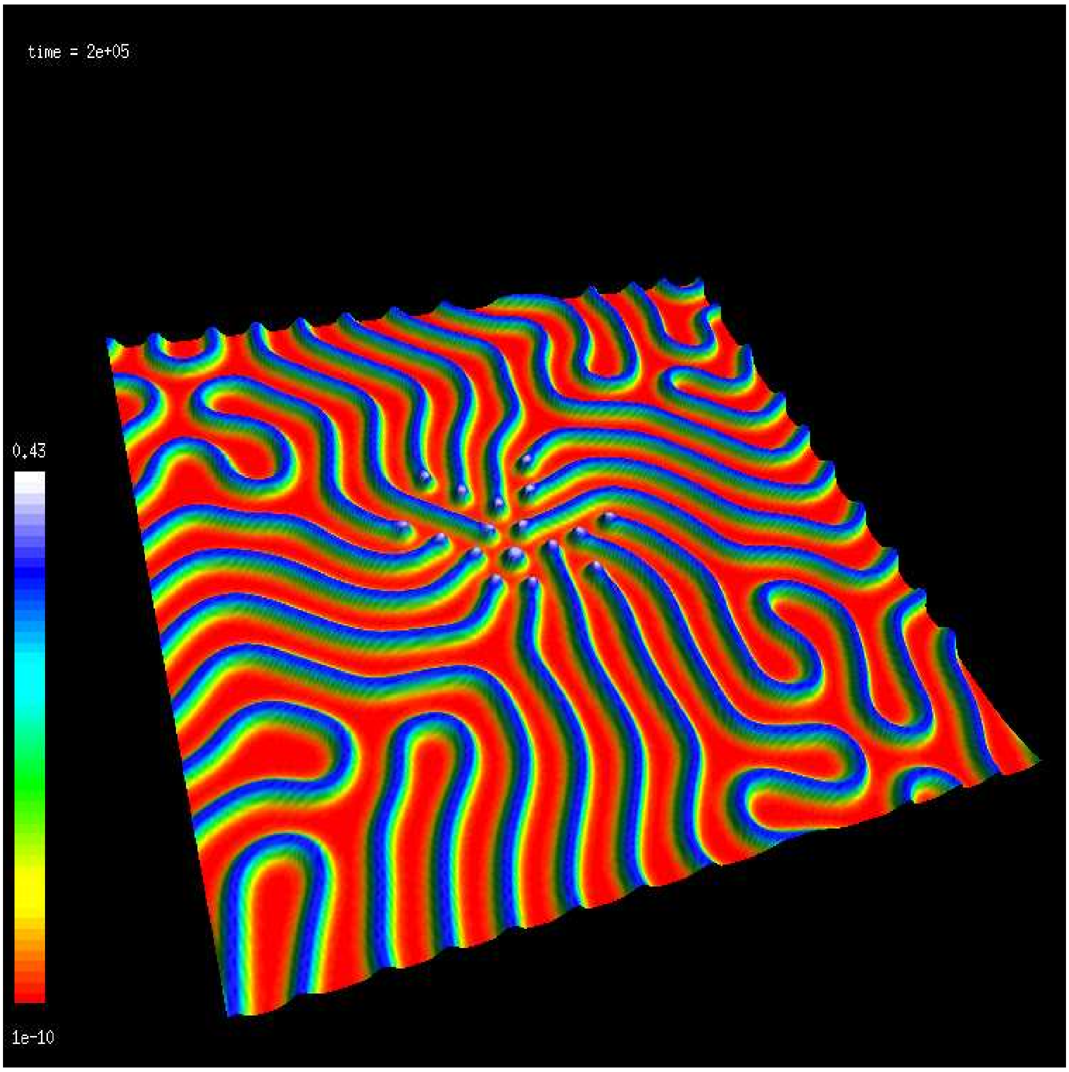


Figure 4.1: The evolved equations of the Gray-Scott model.

As an example of a fully 3D multi-compartment model, the next simulation models a simple signal transduction model of a plasma membrane bound receptor, cytosolic phosphatase and a cytosolic kinase which is activated at the cell surface in a spherical cell originally developed by Brown and coworkers [177]. Extracellular stimulant S reacts with membrane-bound receptor R to produce $S.R$, which in turn phosphorylates kinase K to $K*$ at the membrane. The species $K*$ then diffuses inward to react with P inside the cell. After an initial transient stage, $K*$ reaches a steady state of exponentially decreasing radial concentration (Figure 4.2).

The `reactions.input` file for this simulation contains:

```
locationList {
  location extracellular 1  S;
  location imembrane 1  R S.R S.R.K;
  location cytosol 1;
  default cytosol;
}

numberReactionList {
  S + R <> S.R      :: 4.2 0.25;
  S.R + K <> S.R.K  :: 1.2 0.8;
  S.R.K -> K* + S.R :: 0.2;
  K* + P <> K*.P    :: 1.98 25;
  K*.P -> K + P     :: 6;
}
```

The `locationlist` command defines the compartment of each species. The stimulus S exists only in the extracellular region, the receptor and its intermediates are all on the intracellular membrane and all other species are within the cytosol.

The `numberReactionList` block tells `CellSim` to read the contained equations in terms of quantity (in this case micromoles), instead of quantity/volume (micromolar) concentration that is used in `reactionList`. This option is useful when the rate constants are in terms of quantity and not concentration.

The compartments for this simulation consist of a circle of cytosol with a membrane region overlapping the outermost edge of the cytosol region. At the edges of the cytosolic region, the grid input file for this simulation reads as follows:

```
grid 81 81 81;
...
```

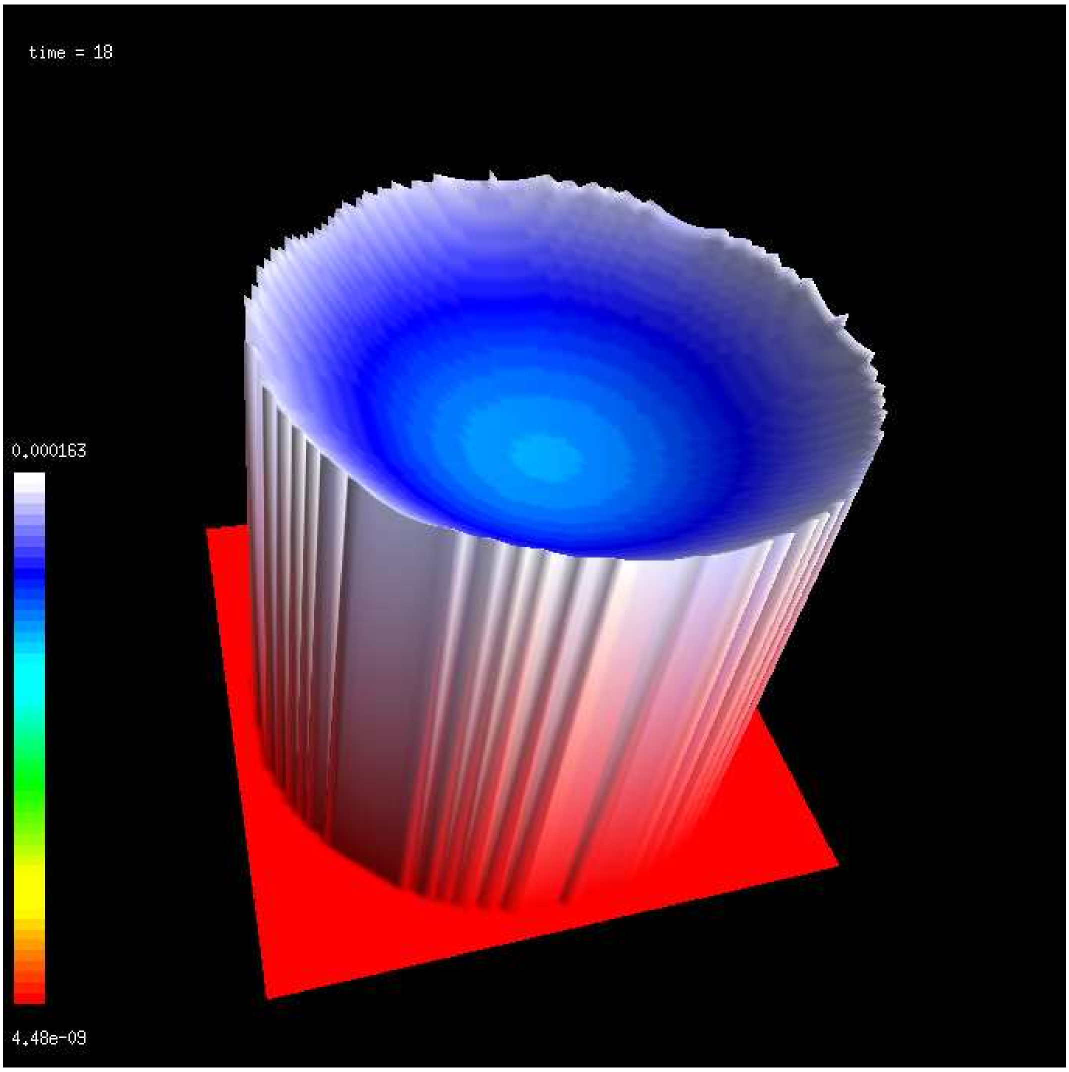


Figure 4.2: `CellSimVis plot` of Cellsim data, showing a 2D slice through the center of a simple 3D signal transduction model. The Z axis represents predicted concentration of a single cytosolic kinase.

```
40  40  2  extracellular;
40  40  3  imembrane cytosol extracellular;
40  40  4  cytosol;
...

80  80  80 extracellular;
```

In the overlapping region the cytosol, membrane, and extracellular compartments all coexist. In this region, reactions involving the stimulus and receptor will occur as well as reactions involving the stimulated receptor and the cytosolic kinase K. As an example of the adaptive integrators in `CellSim`, this simulation uses a second-order Strang-split reaction-diffusion integration scheme:

```
useReactions reactions.input;
useGrid grid.input;
useInitConcentrations initialconc.input;

printOutput output1 {
  printinfo 10;

  printplane 10 plot/K.plane.plot  K   40 * *;
  printplane 10 plot/K*.plane.plot K*  40 * *;

  printgrid 10 plot/K*.grid.plot K*;
  printgrid 10 plot/K.grid.plot K;
}

integrate arb2 {

  dtguess = 1e-2;
  dt = 1.0;
  dx = 0.009765625;
  runtime = 100;

  useStrang;
  runDiffusion;
  diffusionTolerance 1e-2;

  safety = 0.9;
  tolerance = 1e-2;
  dtmin = 1e-10;

  use output1;
}
```

The new print command `printplane` prints the plane specified by the x, y, z coordinates following the filename, where the integer coordinate specifies the position of the plane through the grid and the asterisks define the direction the plane faces. In this example, the simulation prints out the plane of grid points defined by the equation $x = 40$.

The integrate command `integrate arb2` instructs `CellSim` to integrate the system equations using a second-order adaptive Rosenbrock method. The previously unseen commands within the integrate block are specific to the adaptive integrator:

`dtguess`

The initial time step for the adaptive integrator.

`dtmin`

The minimum time step allowed.

`safety`

The maximum increase of a time step is internally set to 50%. This value sets the fraction (0-1) of the maximum increase that should be used.

`tolerance`

Directs the adaptive method to choose the maximum time step that still achieves a given relative accuracy for the kinetics calculation. In this example, a relative accuracy of 0.01 is required.

`diffusionTolerance`

Similar to the tolerance command, this command directs the adaptive step-doubling diffusion calculation to achieve the given relative accuracy.

Sensitivity Analysis

Consider the system defined previously (Equation 4.30). Suppose one would like to calculate $\frac{\partial C_1}{\partial k_1}$ and $\frac{\partial C_4}{\partial k_5}$. The reaction file for this system is defined in the same format as before:

```
locationList {
     location cytosol 1;
     default cytosol;
}
volumeToLiters = 1.0;
reactionList {
   C1 + C1 + C2 <> C3 :: 1e-5 2e-5 k1 k2;
   C4 <> C5 :: 1E-2 1E-3 k3 k4;
}
```

This includes the addition of optional labels $k1, ..., k4$ appended to each reaction. `CellSim` has the ability to use the same label as part of multiple reactions in addition to the notation above. This is sometimes useful for parameter optimization in which several parameters are tied together and optimized as a single identity. Similarly, the initial concentration file has the same format as before:

```
C1 = 100;
C2 = 10;
C3 = 5;
C4 = 1;
C5 = 5;
```

In this example a purely kinetic model is being used, so our grid file consists only of the following:

```
grid 1 1 1;
0 0 0 cytosol;
```

Finally, the main scripting file must carry new instructions to define, calculate, and print the sensitivity parameters.

```
use Reactions reactions.input;
useGrid grid.input;
useInitConcentrations initconc.input;

defineAnalyticalDerivative {
   numerator = C1 C5;
   denominator = k1 k4;
}

printOutput output1 {
    printkinetics 10 screen C1;
    printsensitivity 1 screen C1 k1;
    printsensitivity 1 screen C5 k4;
    printsensitivity 1 dc1dk1 C1 k1;
    outputappendstring = analyticalDerivative;
    outputprependstirng = plot/;

}

integrate arb4_sa {
  use output1;
  dt = 1E-5;
  safety = 0.9;
  dtmin = 1E-10;
  dtguess = 1E-5;
```

```
  tolerance = 1E-4;
}

runAnalyticalDerivative 100;
```

The new commands not previously seen begin with the `defineAnalyticalDerivative` command. Two required subcommands are `numerator` and `denominator`. The numerator must be followed by a list of defined species, while the denominator may be either a list of species or labeled rate constants. If species are used, sensitivity analysis with regards to the initial concentration of that species is carried out. All combinations of derivatives of the numerator and denominator are analytically evaluated throughout the sensitivity run.

The `outputappendstring` command optionally appends any printed files of derivatives by the argument string. Similarly, `outputprependstring` prepends the filename. In the example file, only a single derivative is being printed to a file whose name will be `plot/dC1dk1.analyticalderivative`. The `printOutput` command has a single new command named `printsensitivity`. Its format is similar to that of `printgrid` except that two parameters (`numerator` and `denominator`) are used to define the derivative to be printed.

In this case, the two derivatives $\frac{\partial C_1}{\partial k_1}$ and $\frac{\partial C_5}{\partial k_4}$ are printed to the screen and the former is also printed to a file. The next step defines a Rosenbrock integrator designed especially for sensitivity analysis. This new definition makes sure that enough memory is allocated for the extended Jacobian J_C. Finally, the `runAnalyticalDerivative` tells `CellSim` to perform the calculation for 500 units of time.

4.5.5 `CellSim` Visualization

The visualization component of `CellSim` is a separate program, modeled on a client-server package, and used as a simple monitoring tool for large jobs as well as for visualization of generated data. `CellSimVis` is based on the freely available GPL Licensed QT widget set from Trolltech, (used to develop the popular Unix environment KDE on the GNU/Linux platform). To compile the GUI, OpenGL, QT, and the QT development libraries must be installed. `CellSimVis` may be downloaded and compiled as follows:

Step 1: Login to the anonymous cvs server as before using:

```
export CVSROOT = :pserver:anonymous@bio.physics.drexel.edu:/user/local/cvs-repository
cvs login
```

Step 2: Check out the source code for the GUI with the command:

```
cvs co cellsim-gui
```

Step 3: Then run the following commands using:

```
qmake
make
```

The current version of `CellSim`Vis has the following features:

1. Exportation of rendered images for publication (PNG format)
2. Exportation of MPEG movie files
3. Import of `CellSim` plot files
4. Socket based monitoring of `CellSim` and interactive switching of exported species from `CellSim`
5. OpenGL-based 3D contour plot rendered as 2D grids (solid surface, line mesh, or points)
6. Rendered visualization of surface normals
7. Automatic scaling of model

Imaging of multidimensional data is available within `CellSimVis`. Simple plotting (concentration versus time) is hardware rendered using OpenGL primitives. For two-dimensional data (versus time), visualization is implemented both as a simple 2D color contour plot and as rendered 3D plot of the data with the height representing the concentration on the plane (rendered in real time as a movie). For reading from saved data files, a slider is available for data examination that enables the user to track the changes in concentration with time.

Plot files generated by `CellSim` may easily be imported from the menu (File → Open). For socket based communication, (File → Sockets) should be used. Specify the hostname and socket for the machine running `CellSim`. In this mode, `CellSim`Vis will automatically import all available species in the simulation and make them available for rendering. Snapshots of the visualization can be saved as PNG files using (File → Save).

4.6 Spatio-Temporal Imaging

In recent years, high-resolution single cell imaging has been recognized as a most favorable way to look at biology [217]. Cytomics aims to provide cellular information by executing imaging in a high-throughput, high-content fashion. This information can be used to classify cells, identify molecular hot spots and carry out statistical correlation across different levels of biological hierarchy. Cytomics approaches can also be applied in functional genomics research to characterize the location of proteins [218] and subcellular phenotypes specific to RNAi knockouts in high-throughput assays [219].

These screening technologies provide end-points for a precise description and classification of cells and subcellular phenotypes and a framework for spatio-temporal systems biology. As an example, basic morphological properties of cells have been used to increase the realism of computational models [39]. However, in view of the complex cellular machineries being investigated, the goal would be to perform both a time-resolved and multiplexed analysis at high 3D resolution, within spatially distinguishable compartments inside single cells. Fluorescence confocal microscopy is ideal to perform these tasks.

While confocal microscopes come in a variety flavors, they serve a common function: performing non-invasive optical sectioning using low intensity light (and therefore low radiation damage). This is ideal for studying structural and functional properties of living cells at full 3D microscopic resolution. Specific experimental perturbations can be introduced and monitored, and the quantified cellular behavior can be used to classify distinct cellular phenotypes or phenomes (see chapter 4 of [8]). However, tagging cellular structures and species with multiple fluorescent dyes are limiting factors, although a wealth of non-invasive fluorescent probes has been developed and is now available for monitoring membranes and cell compartments as well as specific protein targets in living specimens [220].

As a specific example, for the study of signaling pathways one would require quantification of the

(i) localization, (ii) concentration, (iii) dynamics, (iv) interaction and (v) activation (phosphorylation) status of many components involved in the cascade. Newly developed fluorescent technologies move this field forward, such as quantum dot (Q-dot) nanoparticles that are ideal to image both localization and concentration of target proteins [221]. In a study of EGFr internalization, an average of 30000 internalized receptors could be monitored at the single cell level, and subsequent image analysis provided regional, average information about concentration and active endosomatic transport.

A distinct advantage of Q-dot assays is their capability to be multiplexed, however they are currently limited in their ability to provide protein interactions and status of protein activation. Diffusion processes can be measured by fluorescence recovery after photobleaching (FRAP) and fluorescence correlation spectroscopy (FCS) [222]. GFP fusion proteins are ideal for FRAP, since they can be bleached without detectable damage to cells. With these tools, differences in the diffusion constant D due to membrane association, scaffolding and compartmentalization can be measured.

In order to detect protein associations in the $1 - 10nm$ range, fluorescence resonance energy transfer (FRET) is the preferred imaging technique [223]. In conjunction with radiometric sensors like EGFR-ECFP and PTB-EYFP in one molecule, FRET can be used to monitor phosphorylation dynamics [224]. Both FRAP and FRET related technologies are currently limited in monitoring multiple species simultaneously. As these fields progress, they will determine the realism of comprehensive spatio-temporal models of regulation in signaling networks, nuclear processes and morphogenesis.

4.7 Conclusion

A biological cell is a complex environment for chemical reactions with a vast and diverse collection of active and passive transport mechanisms, membrane surfaces, and compartments. A new generation of microscopic imaging techniques capable of real-time tracking of single molecules in living cells provides visible evidence of the biological significance of processes dynamically evolving in both space and time. Computer simulation of physics-based models, coupled with quantitative spatio-temporal data will allow cell biologists to rigorously develop and test complex hypotheses. While

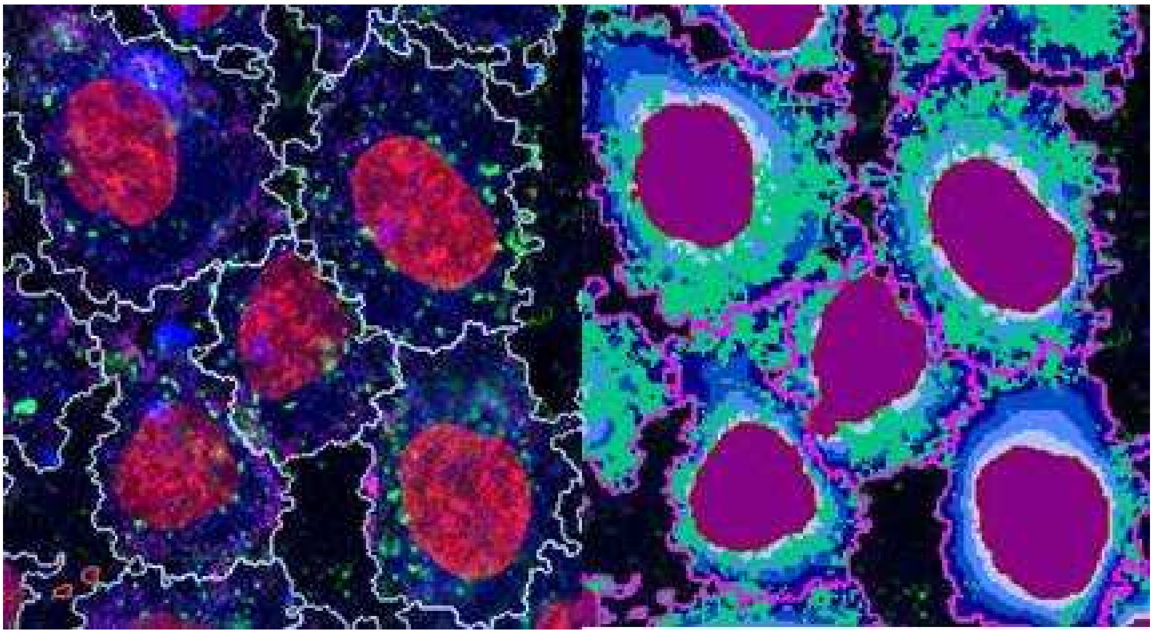


Figure 4.3: Computational imaging delivers quantitative description of the internalization of EGFR, activated by a biotinylated EGF/streptavidin quantum dot complex (green) with A431 cells. Transport routes of internalized EGFR can be monitored by *in vivo* imaging, as well as concentration increase over the time of the experiment (left to right). Concentrations of q-dots within equidistant zones of the cytosol (right) of many cells deliver average information of dynamical processes that feed spatio-temporal systems biology.

methods to simulate reaction-diffusion systems have been successfully applied to complex physical systems such as the atmosphere, oceans, and engine combustion, cells present an unprecedented complexity of significant molecular species, transport mechanisms, and a continuing challenge for experimental measurement.

The relatively small size of the cell also presents a challenge, as many relevant processes occur on atomistic scales that are unsuitable for the continuous, deterministic approach described in this chapter. However, remarkably, cell imaging data suggests that a variety of cell processes are amenable to a reaction-transport model, and the number of proteins per cell generally range from several hundred to hundreds of thousands of each species, supporting the use of molecular concentrations. To address biological problems for which discrete, stochastic approaches are more suitable, several stochastic simulation methods have been proposed (reviewed in [204]). Regardless of the algorithm used, it is necessary to develop tools to interpret simulation results, including efficient sensitivity analysis and interactive, simple interfaces. The emergence of quantitative techniques in cell biology is ushering an era of “predictive” biology and medicine, in which experiments and computer simulation will be blended to help study disease mechanisms and identify therapeutic targets.

Chapter 5: A Fully Adaptive Reaction-Diffusion Integration Scheme with Applications to Systems Biology

5.1 Introduction

Fully-adaptive time integration for multi-dimensional reaction-diffusion PDE problems requires more than a simple error estimate from each integration operator at each time step. Truncation error can also arise from the method in which the operators are applied. A popular second order method for combining operators, particularly in atmospheric science [225], is Strang's method of operator splitting or "Strang splitting" [205, 226, 227]. In the first part of this paper, we derive an expression for the truncation error resulting from Strang splitting reaction and diffusion operators to form a second order integration scheme. We then present explanations of the specific implicit reaction and implicit diffusion operators used as well as their individual error calculations. The method is implemented in CellSim [8, 2], a PDE-based cell simulation software package developed by our group and freely available under the GNU Public License [212]. Finally, we give examples of and discuss the use of this calculation for problems in systems biology.

5.2 Strang Splitting Truncation Error

We first derive an expression for the truncation error due to Strang splitting reaction and diffusion operators [225]. Our generalized system consists of a vector $\boldsymbol{C}(\vec{x},t)$ of chemical concentrations that evolve in time and space according to specified differential equations

$$\frac{\partial}{\partial t}\boldsymbol{C}(\vec{x},t) = F(\boldsymbol{C}(\vec{x},t)) = F_R(\boldsymbol{C}(\vec{x},t)) + F_D(\boldsymbol{C}(\vec{x},t)) \tag{5.1}$$

where the expressions

$$F_R(\boldsymbol{C}(\vec{x},t)) = R(\boldsymbol{C}(\vec{x},t)) = \sum_{i=0}^{n}\left(\alpha_i \prod_{j=0}^{m}[\boldsymbol{C}_j(\vec{x},t)]^{\beta_{ij}}\right),\ \alpha_i \in \mathbb{R}, \beta_{ij} \in \mathbb{N}_0 \tag{5.2}$$

$$F_D(\boldsymbol{C}(\vec{x},t)) = \boldsymbol{D}\nabla^2 \boldsymbol{C}(\vec{x},t) \tag{5.3}$$

are the source term vectors for chemical kinetics and for simple constant-rate diffusion, respectively. The kinetics are composed of a sum over n terms, each of which is a product of a rate constant α_i and concentrations from amongst the system's m reactants, and $\boldsymbol{D}$ is a diagonal matrix of diffusion constants. The exponent β_{ij} determines which reactants contribute to each term. Throughout the paper, we will assume that all operators possess both time and space dependence, and we will specify this dependence when required for clarity. The function space $\mathbf{S}$ consists of all operators of interest that can act on $\boldsymbol{C}$. For reaction-diffusion systems, this consists of $\{I, F_R, F_D\}$ and any linear combination of these.

Following the work of Lanser and Verwer [225], using Lie operator notation [228] adapted from Sanz-Serna and Calvo [229, 230], we start with a reaction-diffusion solution operator $\mathcal{S}(\delta t)$ for Equation (5.1) that acts on a solution at time t to give a solution at time $t + \delta t$,

$$\boldsymbol{C}(t + \delta t) = \mathcal{S}(\delta t)\boldsymbol{C}(t). \tag{5.4}$$

Let $\tilde{\mathcal{S}}(\delta t)$ denote a numerical approximation to $\mathcal{S}(\delta t)$, such that $\tilde{\boldsymbol{C}}(t + \delta t) = \tilde{\mathcal{S}}(\delta t)\boldsymbol{C}(t)$.

By combining the reaction and diffusion sub-operators using Strang splitting [205] and using the preferred order of the stiff and non-stiff operators as outlined in [206, 231] and further applied in [232, 207], the solution operator becomes

$$\tilde{\mathcal{S}}(\delta t) \equiv \tilde{\mathcal{S}}_R(\frac{\delta t}{2})\tilde{\mathcal{S}}_D(\delta t)\tilde{\mathcal{S}}_R(\frac{\delta t}{2}). \tag{5.5}$$

By this, we mean the following:

1. Propagate the system concentrations forward a half step in time using the reaction operator.

2. Propagate the resulting concentrations forward a full step in time using the diffusion operator.
3. Propagate the resulting concentrations forward a final half step in time using the reaction operator.

The evolution of the concentrations can be written as the result of operating on the concentrations with an operator $\hat{F}$

$$\frac{\partial \boldsymbol{C}(t)}{\partial t} = \hat{F}\,\boldsymbol{C}(t), \tag{5.6}$$

and a solution can be found using the solution operator

$$\boldsymbol{C}(t+\delta t) = e^{\hat{F}\delta t}\ \boldsymbol{C}(t). \tag{5.7}$$

The truncation error associated with using the approximate solution operator from Equation (5.5) can be found as follows: For each operator F_i in our function space $\mathbf{S}$, a Lie operator $\mathcal{F}_i$ is associated. This linear operator $\mathcal{F}_i$ maps *any* operator G in $\mathbf{S}$ into the operator $\mathcal{F}_i \cdot G$ such that [229]

$$(\mathcal{F}_i \cdot G)(\boldsymbol{C}) = F_i(\boldsymbol{C})\frac{\partial}{\partial \boldsymbol{C}}G(\boldsymbol{C}). \tag{5.8}$$

We will use this property to find a propagator for the problem at hand. Let F_i be the operator $F = \frac{\partial}{\partial t}$ defined by the left hand side of Equation (5.1). Substituting F into the mapping of Equation (5.8) yields

$$\begin{aligned} (\mathcal{F} \cdot G)(\boldsymbol{C}) &= \frac{\partial \boldsymbol{C}}{\partial t}\frac{\partial}{\partial \boldsymbol{C}}G(\boldsymbol{C}) && (5.9)\\ &= \frac{\partial}{\partial t}G(\boldsymbol{C}). && (5.10) \end{aligned}$$

Moreover, by recursively applying Equation (5.8) using F we find

$$(\mathcal{F}^k \cdot G)(\boldsymbol{C}) = \frac{\partial^k}{\partial t^k} G(\boldsymbol{C}). \tag{5.11}$$

Evaluating these derivatives at time $t = 0$, we find the Taylor expansion of G about time t,

$$\left(\sum \frac{\delta t^k \mathcal{F}^k}{k!} \cdot G\right) \boldsymbol{C}(t) = \left(\left(e^{\delta t \mathcal{F}}\right) \cdot G\right) \boldsymbol{C}(t). \tag{5.12}$$

Of particular interest is the case of $G = I$, the identity operator. Substituting it in reveals

$$\begin{aligned}
\left(\left(e^{\delta t \mathcal{F}}\right) \cdot I\right) \boldsymbol{C}(t) &= \left(I + \delta t \mathcal{F} I + \frac{\delta t^2}{2} \mathcal{F}^2 I + \dots\right) \boldsymbol{C}(t) &(5.13)\\
&= \left(1 + \delta t \frac{\partial}{\partial t} + \frac{\delta t^2}{2} \frac{\partial^2}{\partial t^2} + \dots\right) \boldsymbol{C}(t) &(5.14)\\
&= \boldsymbol{C}(t + \delta t), &(5.15)
\end{aligned}$$

thus propagating the system a step δt forward in time. It is critical to point out that the above procedure spanning Equations (5.9-5.15) applies for all operators in $\mathbf{S}$, not just $F = \frac{\partial}{\partial t}$. Consequently, as our Strang-split propagation operator is actually a combination of sub-operators, this treatment applies such that

$$e^{\delta t \tilde{\mathcal{F}}} \equiv e^{\frac{\delta t}{2} \mathcal{F}_R} e^{\delta t \mathcal{F}_D} e^{\frac{\delta t}{2} \mathcal{F}_R}, \tag{5.16}$$

where we have performed the above procedure three times, using the appropriate ordering of $\mathcal{F}_R$ and $\mathcal{F}_D$, the Lie operator representations of our reaction and diffusion operators from the splitting scheme in Equation (5.5). We note that $\tilde{\mathcal{F}}$ represents a numerical approximation to $\mathcal{F}$.

Since the reaction and diffusion operators do not in general commute, we make use of the BCH formula [233, 234, 235] for gauging the splitting error. The BCH formula stipulates that an expression of linear operators X and Y of the form $e^X e^Y$ can be written as e^Z in terms of commutators, where

$$\begin{aligned} Z = X \quad &+ \quad Y + \frac{1}{2}[X,Y] \\ &+ \quad \frac{1}{12}([X,[X,Y]] + [Y,[Y,X]]) + \frac{1}{24}[X,[Y,[Y,X]]] + \ldots . \end{aligned} \tag{5.17}$$

Applying this formula twice to Equation (5.16) leads to an expression for the exponent:

$$\tilde{\mathcal{F}} = \mathcal{F}_D + \mathcal{F}_R + \frac{\delta t^2}{24}[\mathcal{F}_R,[\mathcal{F}_D,\mathcal{F}_R]] + \frac{\delta t^2}{12}[\mathcal{F}_D,[\mathcal{F}_D,\mathcal{F}_R]] + \mathcal{O}(\delta t^4), \tag{5.18}$$

where we note that odd-ordered terms cancel. The second order and higher terms of Equation (5.18) constitute the error due to splitting the exact reaction and diffusion Lie operators $\mathcal{F}_R$ and $\mathcal{F}_D$. What we are interested in finding is a similar expression to Equation (5.18) but in terms of the numerical operators F_R and F_D,

$$\frac{\tilde{\partial}}{\partial t}\boldsymbol{C}(t) = \tilde{F}(\boldsymbol{C}) \equiv F_R(\boldsymbol{C}) + F_D(\boldsymbol{C}) + \delta t^2 E_F(\boldsymbol{C}) + \mathcal{O}(\delta t^4), \tag{5.19}$$

thus E_F will reveal the second order splitting error of the propagation operator from Equation (5.16). To find this we must first convert E_F from an expression of Lie operators to an expression of numerical operators in $\mathbf{S}$.

Equation (5.8) reveals that a commutator of Lie operators $[\mathcal{F}_1,\mathcal{F}_2]$ is itself a Lie operator associated with a Lie-Poisson bracket of F_1 and F_2,

$$\begin{aligned} [\mathcal{F}_1,\mathcal{F}_2]I(\boldsymbol{C}) \quad &= \quad F_1(\boldsymbol{C})\frac{\partial}{\partial \boldsymbol{C}}\left(F_2(\boldsymbol{C})\frac{\partial}{\partial \boldsymbol{C}}I(\boldsymbol{C})\right) \\ &\qquad\quad -F_2(\boldsymbol{C})\frac{\partial}{\partial \boldsymbol{C}}\left(F_1(\boldsymbol{C})\frac{\partial}{\partial \boldsymbol{C}}I(\boldsymbol{C})\right) \qquad (5.20) \\ &= \quad F_1(\boldsymbol{C})\frac{\partial F_2}{\partial \boldsymbol{C}}(\boldsymbol{C}) - F_2(\boldsymbol{C})\frac{\partial F_1}{\partial \boldsymbol{C}}(\boldsymbol{C}) \qquad (5.21) \\ &= \quad \{F_1(\boldsymbol{C}),F_2(\boldsymbol{C})\}\,. \qquad (5.22) \end{aligned}$$

Expanding the nested commutators that appear in Equation (5.18) reveals their form for numerical operators,

$$\begin{aligned}[\mathcal{F}_1,[\mathcal{F}_2,\mathcal{F}_3]] \quad & I(\boldsymbol{C}) = \\ & \{F_2(\boldsymbol{C}),F_3(\boldsymbol{C})\}^{'}F_1(\boldsymbol{C}) - F_1(\boldsymbol{C})^{'}\{F_2(\boldsymbol{C}),F_3(\boldsymbol{C})\}, \end{aligned} \tag{5.23}$$

where the primes signify partial differentiation with respect to $\boldsymbol{C}$.

To find the term E_F, replace the two nested commutators in Equation (5.18) with their corresponding expressions of Equation (5.23). Doing so reveals

$$\begin{aligned}\frac{\tilde{\partial}}{\partial t} \quad & \boldsymbol{C}(t) = \tilde{F}(\boldsymbol{C}) \equiv F_R(\boldsymbol{C}) + F_D(\boldsymbol{C}) \\ & +\delta t^2 \frac{1}{24}\left[F_{DR}^{'}(F_R + 2F_D) - (F_R^{'} + 2F_D^{'})F_{DR}\right] + \mathcal{O}(\delta t^4), \end{aligned} \tag{5.24}$$

where

$$F_{DR} = F_R^{'}F_D - F_D^{'}F_R \tag{5.25}$$

is the Lie-Poisson bracket from Equation (5.20). Equation (5.24) contains the original components of Equation (5.1) plus terms describing the splitting error and leads to the following expression for E_F:

$$E_F = \frac{1}{24}\left[F_{DR}^{'}(F_R + 2F_D) - (F_R^{'} + 2F_D^{'})F_{DR}\right]. \tag{5.26}$$

Substituting into this using Equations (5.2), (5.3), and (5.25), we arrive at

$$\begin{aligned} E_F \quad &= \frac{1}{24}\Bigg[\left(R^{'}(\boldsymbol{C})\boldsymbol{D}\nabla^2(\boldsymbol{C}) - \left(\boldsymbol{D}\nabla^2(\boldsymbol{C})\right)^{'} R(\boldsymbol{C})\right)^{'} \left(R(\boldsymbol{C}) + 2\boldsymbol{D}\nabla^2(\boldsymbol{C})\right) \\ &\quad - \left(R(\boldsymbol{C}) + 2\boldsymbol{D}\nabla^2(\boldsymbol{C})\right)^{'} \left(R^{'}(\boldsymbol{C})\boldsymbol{D}\nabla^2(\boldsymbol{C}) - \left(\boldsymbol{D}\nabla^2(\boldsymbol{C})\right)^{'} R(\boldsymbol{C})\right)\Bigg] \end{aligned} \tag{5.27}$$

For all *linear* operators $F_l \in \mathbf{S}$ with *any* operator $F_i \in \mathbf{S}$ the following property holds:

$$F_l^{'}(\boldsymbol{C})F_i(\boldsymbol{C}) \equiv F_l\left(F_i(\boldsymbol{C})\right). \tag{5.28}$$

Applying this to Equation (5.27), the term $\left(\boldsymbol{D}\nabla^2(\boldsymbol{C})\right)^{'} R(\boldsymbol{C})$ becomes $\boldsymbol{D}\nabla^2(R(\boldsymbol{C}))$. Expanding out the braced expression leads to

$$\boldsymbol{D}R^{'}(\boldsymbol{C}) \qquad \nabla^2\boldsymbol{C} - \boldsymbol{D}(\nabla^2(\boldsymbol{C}))^{'}R(\boldsymbol{C})$$

$$\begin{aligned} &= \boldsymbol{D}R^{'} \cdot (\boldsymbol{C}_{xx} + \boldsymbol{C}_{yy} + \boldsymbol{C}_{zz}) \\ &\quad -\boldsymbol{D}\vec{\nabla} \cdot \left(\hat{i}(R^{'}\boldsymbol{C}_x + R_x) + \hat{j}(R^{'}\boldsymbol{C}_y + R_y) + \hat{k}(R^{'}\boldsymbol{C}_z + R_z)\right) \end{aligned} \tag{5.29}$$

$$\begin{aligned} &= \boldsymbol{D}R^{'} \cdot (\boldsymbol{C}_{xx} + \boldsymbol{C}_{yy} + \boldsymbol{C}_{zz}) - \boldsymbol{D}R^{'} \cdot (\boldsymbol{C}_{xx} + \boldsymbol{C}_{yy} + \boldsymbol{C}_{zz}) \\ &\quad -\boldsymbol{D}R^{''} \cdot (\boldsymbol{C}_x\boldsymbol{C}_x + \boldsymbol{C}_y\boldsymbol{C}_y + \boldsymbol{C}_z\boldsymbol{C}_z) \\ &\quad -2\boldsymbol{D}(R^{'}_x\boldsymbol{C}_x + R^{'}_y\boldsymbol{C}_y + R^{'}_z\boldsymbol{C}_z) - \boldsymbol{D}(R_{xx} + R_{yy} + R_{zz}), \end{aligned} \tag{5.30}$$

where for simplicity, we no longer explicitly write R's dependency on $\boldsymbol{C}$. After canceling terms, Equation (5.29) can be incorporated into the expanded expression for Equation (5.26).

$$\begin{aligned} E_F = \quad & \frac{-\boldsymbol{D}}{24}\Big[\Big(R^{''} \cdot (\boldsymbol{C}_x\boldsymbol{C}_x + \boldsymbol{C}_y\boldsymbol{C}_y + \boldsymbol{C}_z\boldsymbol{C}_z) \\ & +2(R^{'}_x\boldsymbol{C}_x + R^{'}_y\boldsymbol{C}_y + R^{'}_z\boldsymbol{C}_z) + (R_{xx} + R_{yy} + R_{zz})\Big)^{'}\left(R + 2\boldsymbol{D}\nabla^2(\boldsymbol{C})\right) \\ & -\Big(R + \underline{2\boldsymbol{D}\nabla^2(\boldsymbol{C})}\Big)^{'}\Big(R^{''} \cdot (\boldsymbol{C}_x\boldsymbol{C}_x + \boldsymbol{C}_y\boldsymbol{C}_y + \boldsymbol{C}_z\boldsymbol{C}_z) \\ & +2(R^{'}_x\boldsymbol{C}_x + R^{'}_y\boldsymbol{C}_y + R^{'}_z\boldsymbol{C}_z) + (R_{xx} + R_{yy} + R_{zz})\Big)\Big]. \end{aligned} \tag{5.31}$$

The terms $\boldsymbol{C}_x$, R_{xx}, *etc.*, represent first and second spatial derivatives taken in the direction of the subscripts. Expanding the first line of Equation (5.31) and applying Equation (5.28) for the underlined term in Equation (5.31) produces

$$\begin{aligned} E_F = \quad & \frac{-\boldsymbol{D}}{24}\Big[\Big(R^{'''} \cdot (\boldsymbol{C}_x\boldsymbol{C}_x + \boldsymbol{C}_y\boldsymbol{C}_y + \boldsymbol{C}_z\boldsymbol{C}_z) \\ & +2(R^{''}_x\boldsymbol{C}_x + R^{''}_y\boldsymbol{C}_y + R^{''}_z\boldsymbol{C}_z) + (R^{'}_{xx} + R^{'}_{yy} + R^{'}_{zz})\Big)\left(R + 2\boldsymbol{D}\nabla^2\boldsymbol{C}\right) \\ & -\Big(R^{'} + 2\boldsymbol{D}\nabla^2\Big)\Big(R^{''} \cdot (\boldsymbol{C}_x\boldsymbol{C}_x + \boldsymbol{C}_y\boldsymbol{C}_y + \boldsymbol{C}_z\boldsymbol{C}_z) \\ & +2(R^{'}_x\boldsymbol{C}_x + R^{'}_y\boldsymbol{C}_y + R^{'}_z\boldsymbol{C}_z) + (R_{xx} + R_{yy} + R_{zz})\Big)\Big]. \end{aligned} \tag{5.32}$$

Equation (5.32) is the general form for E_F. For typical simulation geometries, most neighboring grid points will contain the same set of compartments and by extension share the same reaction set. For these homogeneous grid areas, the splitting error can be significantly simplified with spatial derivatives of the reaction operators falling out. Accounting for this yields

$$\begin{aligned} E_{F_h} = \quad & \frac{-\boldsymbol{D}}{24}\Big[\Big(R^{'''} \cdot (\boldsymbol{C}_x\boldsymbol{C}_x + \boldsymbol{C}_y\boldsymbol{C}_y + \boldsymbol{C}_z\boldsymbol{C}_z)\Big)\left(R + 2\boldsymbol{D}\nabla^2(\boldsymbol{C})\right) \\ & -\Big(R^{'} + 2\boldsymbol{D}\nabla^2\Big)\Big(R^{''} \cdot (\boldsymbol{C}_x\boldsymbol{C}_x + \boldsymbol{C}_y\boldsymbol{C}_y + \boldsymbol{C}_z\boldsymbol{C}_z)\Big)\Big]. \end{aligned} \tag{5.33}$$

Further reduction of the expression can be achieved for most biological systems as reactions are typically first and second order. In such cases, $R^{''}$ will be a constant and $R^{'''}$ will fall out of the equations entirely, yielding

$$E_{F_{h,2^{nd}}} = \frac{\boldsymbol{D}}{24}\left(R^{'} + 2\boldsymbol{D}\nabla^2\right)\left(R^{''} \cdot \left(\boldsymbol{C}_x\boldsymbol{C}_x + \boldsymbol{C}_y\boldsymbol{C}_y + \boldsymbol{C}_z\boldsymbol{C}_z\right)\right). \tag{5.34}$$

One approach to the evaluation of Equation (5.32) would be to expand out its second line by evaluating ∇^2 across the terms analytically and then numerically calculating the result. It is however more efficient to evaluate the final group of terms at each spatial grid point and subsequently calculate ∇^2 of the values numerically.

As $\delta t^2 E_F$ has units of $\mu M/s$, we take the value of $\delta t^3 E_F$ as our truncation error estimate in units of concentration. It is this error estimate, along with those of the reaction and diffusion operators, that allow for a full estimation of the numerical error for the propagator over a time step δt.

For stiff systems, Press et al. [209] recommend using a relative error above a given threshold concentration value and an absolute error below. We adopt their recommended scaling of $max(|\boldsymbol{C}|, s)$, and so error values presented in this paper should be interpreted as the maximum *non-negligible* error values used to determine system evolution. Therefore, units for truncation error values will not be specified. The value s is determined by the characteristic scale of the system's concentration values, and we set $s = 1$ for all simulations presented in this paper.

As biochemical reactions are rarely more than bimolecular, Equation (5.34) is often the case for systems of equations modeling homogeneous components of biochemical processes. For the purposes of implementation, we have kept the generalized form (Equation (5.32)) and exploit the sparsity of the high order terms for computational efficiency.

For example, in one dimension the vector term $R^{''}\boldsymbol{C}_x\boldsymbol{C}_x$ may be written component-wise as

$$\left(R^{''}\boldsymbol{C}_x\boldsymbol{C}_x\right)^{(i)} = \sum_{j,k} \frac{\partial^2 R^{(i)}}{\partial \boldsymbol{C}^{(j)} \partial \boldsymbol{C}^{(k)}} \boldsymbol{C}_x^{(j)} \boldsymbol{C}_x^{(k)}. \tag{5.35}$$

This term may be evaluated efficiently by only summing over nonzero $\frac{\partial^2 R^{(i)}}{\partial \boldsymbol{C}^{(j)} \partial \boldsymbol{C}^{(k)}}$.

Similarly, one may efficiently evaluate $R^{'''}\boldsymbol{C}_x\boldsymbol{C}_x(R + 2\boldsymbol{D}\nabla^2\boldsymbol{C})$ component-wise using

$$\begin{aligned}&\left(R^{'''}\boldsymbol{C}_x\boldsymbol{C}_x\,(R \quad + \quad 2\boldsymbol{D}\nabla^2\boldsymbol{C})\right)^{(i)} = \\ &\qquad\qquad\sum_{j,k,l}\frac{\partial^3 R^{(i)}}{\partial\boldsymbol{C}^{(j)}\partial\boldsymbol{C}^{(k)}\partial\boldsymbol{C}^{(l)}}\boldsymbol{C}_x^{(j)}\boldsymbol{C}_x^{(k)}\left(R^{(l)} + 2\boldsymbol{D}\nabla^2\boldsymbol{C}^{(l)}\right).\end{aligned} \tag{5.36}$$

This sum is evaluated over terms that are third order or higher, and while it would seem that for a system of 100 reactants Equation (5.36) would require a sum over 100^3 iterations, $R^{'''}$ is generally very sparse, if not entirely empty, reducing its cost of calculation considerably.

Similarly, other reduced terms are component-wise

$$\left(R^{''}\boldsymbol{C}_x\left(R + 2\boldsymbol{D}\nabla^2\boldsymbol{C}\right)\right)^{(i)} = \sum_{j,k}\frac{\partial^2 R^{(i)}}{\partial\boldsymbol{C}^{(j)}\partial\boldsymbol{C}^{(k)}}\boldsymbol{C}_x^{(j)}\left(R^{(k)} + 2\boldsymbol{D}\nabla^2\boldsymbol{C}^{(k)}\right), \tag{5.37}$$

$$\left(R^{'}\boldsymbol{C}_x\right)^{(i)} = \sum_{j}\frac{\partial^2 R^{(i)}}{\partial\boldsymbol{C}^{(j)}}\boldsymbol{C}_x^{(j)}. \tag{5.38}$$

5.3 Boundaries and Inhomogeneity

In biological cell simulations, the full expression of Equation (5.32) is required in regions where cell compartments border or overlap. At these locations, specific elements of R may differ due to the distinct sets of reactions that take place in the various compartments. To determine the partial derivatives of R with respect to x, we cannot simply numerically difference the values of $R(\boldsymbol{C}(\vec{x},t),\vec{x})$ on the grid as this would yield total derivatives. Instead we must determine new reaction sets for spatial derivatives of R at these locations.

Consider a three-point quadratic interpolation of R over the evenly spaced grid points x_0, x_1, and x_2. The one-dimensional interpolation is expressed as follows

$$\begin{aligned}R(\boldsymbol{C}(x),x) &= R(\boldsymbol{C}(x_0),x_0) + (R(\boldsymbol{C}(x_1),x_1) - R(\boldsymbol{C}(x_0),x_0))\left(\frac{x-x_0}{\Delta x}\right) \\ &+ \frac{R(\boldsymbol{C}(x_2),x_2) - 2R(\boldsymbol{C}(x_1),x_1) + R(\boldsymbol{C}(x_0),x_0)}{2}\left(\frac{x-x_0}{\Delta x}\right)\left(\frac{x-x_1}{\Delta x}\right).\end{aligned} \tag{5.39}$$

Taking the first centered partial derivative with respect to x, $\frac{\partial R}{\partial x}\Big|_{x_1}$ yields the expression

$$\frac{\partial R}{\partial x}\bigg|_{x_1} = \frac{R(\boldsymbol{C}(x_1), x_2) - R(\boldsymbol{C}(x_1), x_0)}{2\Delta x}. \tag{5.40}$$

The second derivative is

$$\frac{\partial^2 R}{\partial x^2}\bigg|_{x_1} = \frac{R(\boldsymbol{C}(x_1), x_2) - 2R(\boldsymbol{C}(x_1), x_1) + R(\boldsymbol{C}(x_1), x_0)}{\Delta x^2}. \tag{5.41}$$

The inhomogeneous equations (5.40) and (5.41) can be predetermined and then evaluated in the same manner as the standard reaction operator equations.

5.4 Reaction and Diffusion Truncation Error Estimates

In addition to the splitting error calculation, a fully adaptive integration scheme must also be able to determine error estimates for the reaction and diffusion operators. For completeness, we present brief explanations of how we determine truncation error estimates for the two operators as well as present explanations of the methods themselves.

5.4.1 Evaluation of the Error for the Diffusion Propagator

For propagating the diffusion operator, we use an ADI method introduced by Douglas [210, 236]. The method easily generalizes to problems of arbitrary dimensionality and is well-suited to our intracellular diffusion problem. In three dimensions it may be written as

$$\frac{\boldsymbol{c}^* - \boldsymbol{c}_n}{\Delta t} = \frac{\alpha}{2}\delta_x^2(\boldsymbol{c}^* - \boldsymbol{c}_n) + \alpha\delta_y^2\boldsymbol{c}_n + \alpha\delta_z^2\boldsymbol{c}_n \tag{5.42}$$

$$\frac{\boldsymbol{c}^{**} - \boldsymbol{c}_n}{\Delta t} = \frac{\alpha}{2}\delta_x^2(\boldsymbol{c}^* - \boldsymbol{c}_n) + \frac{\alpha}{2}\delta_y^2(\boldsymbol{c}^{**} - \boldsymbol{c}_n) + \alpha\delta_z^2\boldsymbol{c}_n \tag{5.43}$$

$$\frac{\boldsymbol{c}_{n+1} - \boldsymbol{c}_n}{\Delta t} = \frac{\alpha}{2}\delta_x^2(\boldsymbol{c}^* - \boldsymbol{c}_n) + \frac{\alpha}{2}\delta_y^2(\boldsymbol{c}^{**} - \boldsymbol{c}_n) + \frac{\alpha}{2}\delta_z^2(\boldsymbol{c}_{n+1} - \boldsymbol{c}_n), \tag{5.44}$$

where $\alpha = \frac{\boldsymbol{D}}{\Delta x^2}$ and δ_x^2 is a tridiagonal matrix representing 1D diffusion along a strip of space in the direction of the subscript. The concentration vector $\boldsymbol{c}$ is a vector whose elements are concentrations of a single reactant at successive locations along the grid strip. For a strip in the x direction having periodic boundary conditions, the matrix δ_x^2 is of the form

$$\begin{bmatrix} -2 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & 1 \\ 1 & -2 & 1 & 0 & \cdots & 0 & 0 & 0 & 0 \\ & & & & \ddots & & & & \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 & -2 & 1 \\ 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 & -2 \end{bmatrix}$$

Notice that due to periodic boundaries the upper right and lower left corner matrix elements are non-zero, and so will require cyclic tridiagonal solving methods [209].

We can somewhat simplify the above ADI equations by subtracting Equation (5.42) from Equation (5.43) and Equation (5.43) from Equation (5.44) and then defining $a = \alpha\delta t$. After doing so, we are left with the following reduced equations:

$$(1 - \frac{a}{2}\delta_x^2)\boldsymbol{c}^* = (1 + \frac{a}{2}\delta_x^2 + a\delta_y^2 + a\delta_z^2)\ \boldsymbol{c}_n \tag{5.45}$$

$$(1 - \frac{a}{2}\delta_y^2)\boldsymbol{c}^{**} = \boldsymbol{c}^* - \frac{a}{2}\delta_y^2\boldsymbol{c}_n \tag{5.46}$$

$$(1 - \frac{a}{2}\delta_z^2)\boldsymbol{c}_{n+1} = \boldsymbol{c}^{**} - \frac{a}{2}\delta_z^2\boldsymbol{c}_n, \tag{5.47}$$

or in the generalized form for an N dimensional system,

$$(1 - \frac{a}{2}\delta_1^2)\boldsymbol{c}_1^* = (1 + \frac{a}{2}\delta_1^2 + \sum_{i=2}^{N} a\delta_i^2)\ \boldsymbol{c}_n \tag{5.48}$$

$$(1 - \frac{a}{2}\delta_i^2)\boldsymbol{c}_i^* = \boldsymbol{c}_{i-1}^* - \frac{a}{2}\delta_i^2\boldsymbol{c}_n \quad for\ i = 2...N \tag{5.49}$$

$$\boldsymbol{c}_{n+1} = \boldsymbol{c}_N^*. \tag{5.50}$$

In one dimension, this is simply the Crank-Nicolson differencing scheme [208]. In three dimensions, determining $\boldsymbol{c}_{n+1}$ consists of evaluating three tridiagonal linear equations, each an $\mathcal{O}(n)$

operation. As the left hand side matrices $(1-\frac{a}{2}\delta^2)$ possess strict diagonal dominance, they are necessarily non-singular [237, 238]. In our own implementation in Cellsim, we make generous use of the Gnu Scientific Libraries for quickly solving these tridiagonal and other linear algebra problems. For both zero flux and periodic boundary conditions, no special treatment is required at the boundaries when calculating the intermediate values, $\boldsymbol{c}^*$ and $\boldsymbol{c}^{**}$.

When the grid is uniformly spaced in all three directions, these intermediate values can be eliminated from the set of equations to yield

$$\begin{aligned} a(\delta_x^2+\delta_y^2+\delta_z^2)\frac{(\boldsymbol{c}_{n+1}+\boldsymbol{c}_n)}{2} &= (\boldsymbol{c}_{n+1}-\boldsymbol{c}_n)+ \\ +\frac{a^2}{2\Delta t}(\delta_x^2\delta_y^2+\delta_y^2\delta_z^2+\delta_z^2\delta_x^2)(\boldsymbol{c}_{n+1}-\boldsymbol{c}_n) &- \frac{a^3}{4\Delta t^2}\delta_x^2\delta_y^2\delta_z^2(\boldsymbol{c}_{n+1}-\boldsymbol{c}_n), \end{aligned} \tag{5.51}$$

The first line of the equation is the standard Crank-Nicolson scheme. The last two terms of Equation (5.51) appear entirely due to the arrangement of the ADI scheme, and are taken to be the second order splitting error of this ADI scheme. Note that in two dimensions, the final step of Equation (5.45) is eliminated, leaving $\boldsymbol{c}^{**}$ as the resultant $\boldsymbol{c}_{n+1}$. Also, the splitting error expression is considerably simplified as all terms containing δ_z^2 fall away, with $\frac{a^2}{2\Delta t}\delta_x^2\delta_y^2(\boldsymbol{c}_{n+1}-\boldsymbol{c}_n)$ as the only remaining error term.

To estimate the truncation error of the method, we first notice that the right side of Equation (5.45) contains all the required elements of the first-order FTCS diffusion scheme [209, 239], which in three dimensions appears as

$$\hat{\boldsymbol{c}}_{n+1} = (1+a\delta_x^2+a\delta_y^2+a\delta_z^2)\ \boldsymbol{c}_n. \tag{5.52}$$

Therefore, the second order ADI method contains an embedded first order FTCS method that can be used for a simple truncation error estimate:

$$\Delta_D = \boldsymbol{c}_{n+1} - \hat{\boldsymbol{c}}_{n+1}. \tag{5.53}$$

The maximum valued element of the vector Δ_D is taken as the truncation error estimate for a single time step diffusion propagation.

5.4.2 Evaluation of the Error for the Reaction Propagator

For propagating chemical kinetics, we employ a Rosenbrock method [209]. Such methods have the general form

$$\boldsymbol{C}_{n+1} = \boldsymbol{C}_n + \sum_{i=1}^{s} b_i \boldsymbol{k}_i \tag{5.54}$$

$$\boldsymbol{k}_i = \delta t F_R \left(\boldsymbol{C}_n + \sum_{j=1}^{i-1} a_{ij} \boldsymbol{k}_j \right) + \delta t \boldsymbol{J} \sum_{j=1}^{i} \gamma_{ij} \boldsymbol{k}_j, \tag{5.55}$$

where a_{ij},b_i, and γ_{ij} are constants, and $\boldsymbol{J}$ is the Jacobian matrix of the chemical kinetics whose elements $\boldsymbol{J}_{ij} = \frac{\partial F_R(\boldsymbol{C}_i)}{\partial \boldsymbol{C}_j}$. We employ the second order method developed in [211]. The method is written

$$\begin{aligned} \boldsymbol{C}_{n+1} &= \boldsymbol{C}_n + \frac{3}{2}\delta t \boldsymbol{k}_1 + \frac{1}{2}\delta t \boldsymbol{k}_2, \\ (I - \gamma \delta t \boldsymbol{J})\, \boldsymbol{k}_1 &= F_R(\boldsymbol{C}_n), \\ (I - \gamma \delta t \boldsymbol{J})\, \boldsymbol{k}_2 &= F_R(\boldsymbol{C}_n + \delta t \boldsymbol{k}_1) - 2\boldsymbol{k}_1, \end{aligned} \tag{5.56}$$

where $\gamma_{ij} = \gamma = 1 \pm 1/\sqrt{2}$ is chosen for desired stability properties. Further details can be found in [211]. The method contains an embedded first-order method giving an approximate solution of

$$\hat{\boldsymbol{C}}_{n+1} = \boldsymbol{C}_n + \delta t \boldsymbol{k}_1. \tag{5.57}$$

Again, subtracting the first-order solution, Equation (5.57), from the second order solution, Equation (5.56), yields a truncation error estimate of $\Delta_R \equiv \boldsymbol{C}_{n+1} - \hat{\boldsymbol{C}}_{n+1} = \frac{\delta t}{2}\boldsymbol{k}_1 + \frac{\delta t}{2}\boldsymbol{k}_2$. We use the maximum valued element of the vector Δ_R as the truncation error estimate for a single time step reaction propagation.

5.5 The Integration Scheme

Depending upon the model system being integrated, any of the three sources previously described (reaction, diffusion, and Strang splitting) may contribute to the truncation error. Monitoring all three affords one the most information when adjusting the time step during adaptive integration. If any single source's truncation error is above a maximum tolerance Δ_{max} for that source, the error can be reduced to an acceptable value by shrinking the time step value. Correspondingly, if all three truncation errors are below their tolerances, the time step can be increased to improve efficiency for a given desired accuracy. This is the basic approach of our integration scheme.

First, since the splitting calculation is explicit, the truncation error due to splitting can be determined without the need to take the step first. Accordingly our integration scheme checks the splitting error first. If this error surpasses the splitting tolerance, the time step δt is reduced to the maximum value allowed by the tolerance. This can be done without the need to recalculate the splitting error using the smaller step size because the splitting error is directly proportional to δt^3. So, if we calculate a splitting error value of Δ_{fail} which turns out to be larger than the error tolerance Δ_{max}, we need only reduce the step size by a factor of $(\Delta_{max}/\Delta_{fail})^{\frac{1}{3}}$ to find a time step that will result in a splitting error equal to the tolerance. If, on the other hand, the error is less than the tolerance, no adjustment to δt is made.

After determining the splitting error, the system is propagated forward in time an amount $\delta t/2$ by the reaction operator F_R. If the step fails the tolerance test, the time step size is reduced and the step is re-taken until the tolerance test is passed. If the step passes the tolerance test, the value of the next time step ($\delta t'$) is increased and δt is left unchanged.

After this, the system is propagated forward in time an amount δt by the diffusion operator F_D. If the step fails the tolerance test, the time step size is reduced, the partly taken integration step is abandoned, and the entire step is restarted with the reaction propagation. If the diffusion propagation passes the tolerance test, the system is propagated forward in time a final half time step $\delta t/2$ by F_R. Finally, the value for the next time step is set to the saved value $\delta t'$. A flow chart depicting this algorithm can be found in Figure 5.1.

5.6 Example Systems

We now examine three example systems that individually demonstrate system evolution dominated by truncation error from each of the three sources. For the reaction error-dominated system, we use a simple auto-catalytic system diffusing through a medium with two slowly diffusing reactants A and B that react quickly to form a dimer. For the diffusion error-dominated system, we use a similar system with different time scales for reaction and diffusion constants. For the splitting error-dominated system we use a simple simulation of a kinase/phosphatase cascade, activated by a membrane-bound receptor/stimulus complex on a compartmentalized grid. All simulations are integrated using CellSim [8].

As a check for the accuracy of the method, we also integrate out each system using an explicit 4th-order Runge-Kutta (RK4) method and FTCS diffusion with a small fixed time step ($10^{-5}s$ for the two auto-catalytic systems and $10^{-3}s$ for the kinase cascade system), an accurate but computationally costly method. We plot the maximum concentration difference between the two methods at each time step as cumulative error calculations. Since the time points of the two methods do not identically match, we use simple linear interpolation of the RK4 data to determine appropriate values. We provide these values for the first 100 seconds of each simulation, approaching steady state concentrations in all cases. Computational storage and time constraints prevented us from integrating the RK4 simulations out further.

5.6.1 Diffusion Error-Dominated System

Diffusion error-dominated dynamics are characterized by fast diffusion constants compared to other system rate parameters. In our example of a diffusion error-dominated system, we start with a 100×100 square grid of width 2 microns, with an initial Gaussian distribution of the reactants. The relative concentrations of the reactants do not match their kinetic equilibrium concentrations, and the boundary conditions are periodic. The sole kinetic reaction is

$$A + B \rightleftarrows 2A, \tag{5.58}$$

where the forward rate constant is $0.001(\mu M \cdot s)^{-1}$, the backward rate is $0.0001(\mu M \cdot s)^{-1}$, and diffusion constant for both reactants is $5.0\mu m^2/s$. This leads to the reaction and diffusion operators for this system:

$$F_R = \begin{bmatrix} 0.001[A][B] - 0.0001[A][A] \\ 0.0001[A][A] - 0.001[A][B] \end{bmatrix} \tag{5.59}$$

$$F_D = 5.0\nabla^2 \begin{bmatrix} [A] \\ [B] \end{bmatrix}. \tag{5.60}$$

Initial conditions are

$$A(x,y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{5.61}$$

$$B(x,y) = 2A(x,y), \tag{5.62}$$

where $\sigma = 0.205$, and the origin is at the center of the grid. This 2D system is integrated out from time $t_0 = 0s$ to time $t = 1000s$ using the adaptive integration algorithm presented earlier in this paper (Figure 5.1). The maximum allowed error for each source of error is $\Delta_{max} = 0.001$.

The initial dynamics of the reactants can be seen in the first two rows of Figure 5.2. Shown are the concentration profiles after the first step ($t_1 = 0.000125s$), after the initial Gaussians have diffused to approximately half their initial height ($t_2 = 0.00342s$), and after they have diffused to uniform distributions ($t_3 = 0.0995s$). From this point the uniformly distributed reactants rise/fall to steady-state reaction values. Note that the legend scales are different for all rows in these image collages.

The lower three rows of Figure 5.2 show the maximum error values over the grids for the three sources of error – diffusion, reaction, and Strang splitting. For time $t_1 = 0.000125s$, the diffusion error plot contains three points of interest – a central peak corresponding to the central concentration peaks, an inner ring corresponding to the bottom edge of the concentration Gaussian, and a less pronounced third outer ring corresponding to the area just outside the concentration Gaussian. The central peak of truncation error arises simply due to the central curvature of the concentration distribution. The inner and outer truncation error rings correspond to the areas approaching the edge of the concentration Gaussian – from both the zero gradient outside area and the constant gradient slope of the Gaussian. As the concentration Gaussians spread out, these diffusion error rings move radially as well, as shown for time $t_2 = 0.00342s$. By time $t_3 = 0.0995s$, each Gaussian has spread out enough to completely fill the grid. The outer error ring disappears, and the inner ring diameter continues to grow larger than the grid width, causing the four peaks in the grid corners, constituting a single peak wrapped around the periodic boundary edge.

At this point, the initial Gaussian concentration distributions have diffused to a uniform distribution over the grid. This lack of strong curvature for concentration causes the diffusion error to drop drastically for small time steps. In Figure 5.3, the system diffusion error remains pinned just under the maximum allowed error until around $t = 0.1s$. At this point the reactant concentrations have become uniform over the grid, which causes the diffusion error to drop several orders of magnitude (from 10^{-3} to 10^{-6}). From this point on, all three sources of error are below the maximum allowed error, and so the adaptive algorithm gradually increases the system time step δt. As the errors do not reach the maximum allowed error until the very end of the run as seen in Figure 5.4 , δt is increased exponentially from a time step of $\delta t \approx 0.1$ at time $t = 0.1s$ to a time step of $\delta t \approx 500s$ at time $t = 1000s$ (Figure 5.4). Only with such large time steps does the diffusion error re-approach the maximum allowed error at this later stage of the simulation.

The reaction error and splitting error do not play a major role in the early stages of the system evolution, and remain several orders of magnitude below the system dominating diffusion error up to time $t = 0.1s$, as seen in the bottom two rows of Figure 5.2 and in Figure 5.3. At time $t = 0.1s$ the reaction error does not fall off as does the diffusion and splitting error. This is because the reaction error is most sensitive to kinetic rates and concentration values rather than spatial gradients. The reaction error instead increases exponentially along with δt, though it still remains more than two orders of magnitude smaller than the diffusion error.

The cumulative error for the diffusion dominant system remains low throughout the simulation, ranging between $4.11 \times 10^{-4} \mu M$ and $1.31 \times 10^{-5} \mu M$ until a uniform reactant distribution is reached around $t = 0.1s$, at which point the cumulative error drops significantly (Figure 5.3).

5.6.2 Reaction Error-Dominated System

For a reaction error-dominated example, we start with the same two Gaussian distributions of A and B reacting to form $2A$ in an autocatalytic reaction. The grid geometry and initial concentration distributions also remain the same. The forward and backward kinetic rate constants are changed to $1(\mu M \cdot s)^{-1}$ and $0.1(\mu M \cdot s)^{-1}$, respectively. The diffusion constant for both reactants is set to $0.001 \mu m^2/s$ and $\Delta_{max} = 0.001$. The reaction and diffusion operators for this system become:

$$F_R = \begin{bmatrix} [A][B] - 0.1[A][A] \\ 0.1[A][A] - [A][B] \end{bmatrix} \tag{5.63}$$

$$F_D = 0.001 \nabla^2 \begin{bmatrix} [A] \\ [B] \end{bmatrix} \tag{5.64}$$

The time scale of the reaction operator is now much smaller than the diffusion operator.

In Figure 5.5 we examine a series of reactant concentrations and maximum truncation errors over the grid during the initial stages of system evolution – the early reaction error-dominated period from time $t = 0s$ to time $t \approx 3s$. Figure 5.6 shows the same concentrations and errors during the later, diffusion error-dominated stages of system evolution. As the reaction kinetics reach steady-state values, the reaction error no longer dominates and the system integration evolves at a faster rate that allows the slower diffusion to dominate.

In Figure 5.5, we see that the enzymatic reaction kinetics initially dominate, rapidly converting B into A ($t_1 = 0.001s$ to $t_2 = 0.227s$). This increases the height of A's Gaussian distribution and creates a depression at the center of B's distribution ($t_3 = 0.676s$). At time $t_4 = 2.776s$ in Figure 5.6, the "shoulders" of B's distribution ($t_4 = 2.776s$) are reduced to reach a final Gaussian distribution at kinetic equilibrium ($t_5 = 15.064s$).

Eventually, near time $t_6 = 918.65s$, the reactants have diffused uniformly over the grid to steady-state concentrations, which allows for exponential growth of the time step value. As seen in Figure 5.8, δt grows over the full extent of the simulation from an initial time step of $10^{-3}s$ to final steps of over $1000s$.

As for the truncation errors, during the first second of integration the reaction error plays a larger role in the system evolution than the diffusion error and splitting error, though the latter two rise in tandem (Figure 5.7). At time $t_2 = 0.227s$, the reaction error peaks just under the maximum allowed error and then begins decreasing exponentially in three stages: from $t \approx 0.2s$ to $t \approx 1.5s$ after it has been overtaken by the diffusion error, from $t \approx 1.5s$ to $t \approx 100s$, and from $t \approx 100s$ to $t \approx 200s$. After the diffusion/reaction crossover around time $t \approx 1.5s$, the splitting error falls in tandem with the falling reaction error. Beyond $t \approx 200s$, both the the reaction and splitting error remain at low values approaching the lower limit of double precision.

During the drop in the reaction error and splitting error, the diffusion error dominates from time $t_4 = 2.776s$ to time $t_6 = 918.65s$, with the maximum diffusion error hovering just under the maximum allowed error value (Figure 5.7). After this point, the uniform reactant distributions greatly reduce the diffusion truncation error for small time steps, and so the diffusion error drops two orders of magnitude before the increased time step sizes level the diffusion error magnitude at the end of the simulation. Due to the limited time span of the simulation, truncation error values do not re-approach the maximum allowed value

Similarities can be seen between the time step evolution and truncation error evolution of the reaction and diffusion error-dominated system simulations in their later periods. As seen in Figures 5.3 and 5.7, after the reactants reach uniform distribution, the diffusion error drops by multiple orders of magnitude. This allows δt to grow exponentially until errors re-approach specified limits as seen in Figures 5.4 and 5.8.

The similarities can be further illuminated by comparing Figures 5.2 and 5.6. Differences between the two sets of images can largely be attributed to the "shoulder" of B's distribution in Figure 5.6.

For instance, the additional rings present in the diffusion error of the reaction-dominated system (third row in Figure 5.6) can be attributed to this. Once the reactants reach kinetic equilibrium around $t \approx 15s$, the diffusion error evolves quite similarly to that of the diffusion error-dominated system, only at a slower time scale. Conversely, the reaction and splitting error profiles evolve differently between the two simulations. In the diffusion error-dominated system, the reactants have not reached kinetic equilibrium by the time diffusive equilibrium is reached, so reaction error values do not fall as they do in the reaction error-dominated system. Because the splitting error is affected by both reaction and diffusion terms, it also is dissimilar between the two simulations.

Again, the cumulative error for the reaction dominant system remains low throughout the simulation, ranging between $1.11 \times 10^{-3} \mu M$ and $5 \times 10^{-6} \mu M$ for the first 100 seconds of simulation (Figure 5.7). The initial peak in the cumulative error corresponds to the initial fast reaction-dominated dynamics of the system evolution and is mirrored by a less pronounced peak in the reaction truncation error estimate.

5.6.3 Strang Splitting Error-Dominated System

As an example of a system dominated by Strang splitting error, we model a kinase/phosphatase cascade activated by an extracellular stimulus [177]. In our simulation, an extracellular ligand S binds to a transmembrane receptor R, which activates a cytosolic kinase K near the membrane. This activated kinase K^* is free to diffuse within the cytoplasm and undergo dephosphorylation by a phosphatase P. The system equations are as follows:

Reaction			k_f	k_b
$S + R$	$\rightleftarrows$	$S.R$	4.2	0.25
$S.R + K$	$\rightleftarrows$	$S.R.K$	1.2	0.8
$S.R.K$	$\rightarrow$	$K^* + S.R$	0.2	
$K^* + P$	$\rightleftarrows$	$K^*.P$	1.98	25
$K^*.P$	$\rightarrow$	$K + P$	6	

with appropriate units of micromolar concentration and seconds. The initial concentrations of the reactants are

S	0.001	μM
R	1.6667	μM
K	0.2	μM
P	0.224	μM.

All other concentrations are initially zero, and the stimulus concentration S is held constant. All parameter values are adapted from the kinase cascade in [1].

The simulation is run on a 100×100 point square grid of total width $5\mu m$. We model the cell as a circular region whose radius extends to the edge of the grid. The outer grid points of the circular region form the cell membrane region. The *extracellular* region overlaps with the *membrane* region and extends to the grid edge. The *cytosolic* region also overlaps with the *membrane* region and extend into and fills the cell center. The receptor R and all its complexes exist in (and are free to diffuse among) grid points containing the membrane compartment. The stimulus S exists in extracellular grid points, and reactants K, K^*, $K^*.P$, and P all exist in the cytosol. The diffusion constant for all reactants is $1\mu m^2/s$. This system is integrated from time $t = 0$ out to time $t = 200s$ using our adaptive integrator with a tolerance of $\Delta_{max} = 10^{-5}$ for all three sources of error.

We chose this system not only to more directly show the algorithm's applications to problems encountered in systems biology, but also to demonstrate how splitting errors can dominate for compartmentalized, heterogeneous systems of nonlinear reactions. Such systems often occur for spatiotemporal cell signal transduction simulations, where the system reactants (proteins, small molecules, ions, etc.) are kept separate by compartment borders and cell geometry.

As seen in Figure 5.9, K^* concentration rises to a steady state of $0.0853nM$ at the cell edge which decays exponentially inward to a central concentration of $0.075nM$. Almost immediately ($t_1 = 0.071s$, first column of Figure 5.9), the splitting error dominates the system evolution. The error is most prominent at the interface between the *extracellular* and *membrane* compartments located at the outer edges of the cell. After sufficient accumulation of $S.R$ in the membrane, splitting error increases at the interface between the *membrane* and *cytosolic* compartments located at the internal edges of the cell, eventually overtaking the outer-edge splitting error at time $t_2 = 8.327s$

(Figure 5.9, second column). The splitting error at the outer edge reduces slightly by time $t_3 = 100s$ and the system reaches steady state (Figure 5.9, third column).

In Figure 5.9, the reaction error (third row) quickly reaches a peak value at t_1, which then reduces to the distributions seen at times t_2 and t_3. The diffusion error distribution (second row) mimics that of the splitting error and gradually increases in value up to a maximum of 7.63×10^{-6} at time t_2, just below the allowed tolerance.

As seen in Figure 5.10, for the vast majority of integration time, the splitting error is at the maximum allowed value. At steady state, the diffusion error is three-quarters of this value and the reaction error is five orders of magnitude smaller. Therefore, the splitting error determines the time step evolution plotted in Figure 5.11. During the initial integration steps, the time step size increases exponentially until the splitting error reaches the maximum allowed value at time $t \approx 0.03$. From here the time step size increases more slowly until time $t_2 = 8.327s$. From this time on through the remainder of the simulation, the time step size does not significantly change.

Of the seven reactants present in the simulation, it is inactive K that produces the greatest splitting error throughout the simulation. This is localized at the cell membrane where K partakes in the most reactions, involving $S.R$, $S.R.K$, and $K^{*}.P$ simultaneously. It is interesting to note that the splitting error for the reactants S, R, $S.R$ and $S.R.K$ quickly drop off to negligible values after time $t_2 = 8.327s$.

The discontinuous appearance of the splitting and diffusion error is a result of modeling a circular cell on a square grid lattice. Membrane grid points will have neighboring grid points of varying compartments, depending on the circular membrane's orientation to the square lattice direction. Therefore, the three-point numerical differencing used in the diffusion error and splitting error calculations will be affected by the location on the membrane. Despite the random appearance, the concentration and error profiles all have four-fold symmetry (eight-fold mirror symmetry), as would be expected on a square grid.

The cumulative error for the Strang splitting dominant system is highest for the kinase K, as the total error for both the active and inactive form reach values of $3.75 \times 10^{-7} \mu M$ within the first twelve seconds of the simulation (Figure 5.10).

5.7 Discussion and Final Remarks

We have shown that the evolution of even the simplest reaction-diffusion systems likely to be encountered in multi-dimensional systems biology simulations can be dominated by truncation error due to both reaction and diffusion operators, as well as the truncation error due to the common second-order Strang's method of operator splitting. In multi-dimensional cell signaling simulations, truncation error due to Strang splitting can entirely dominate the system evolution as demonstrated by our third example simulation.

We have presented a method of monitoring the truncation error due to Strang splitting and incorporating it into an adaptive step size integration algorithm. While truncation error and global error monitoring and control of individual operators is a common field of study [240, 241], time adaptive methods that incorporate Strang splitting truncation error as an input to time step control have, to our knowledge, not been previously evaluated.

Operator splitting and time splitting methods have found popularity in a wide variety of simulation applications including Bose-Einstein condensation [242], quantum statistical calculations [243], optical interactions in media [244], and transport in porous media [245]. Moreover, Strang splitting is used extensively in the field of atmospheric simulation. This lead Lanser and Verwer [225] to conjecture that splitting errors are kept within practical bounds for typical problems encountered in the field. As we have shown, this statement cannot be made, in general, for problems in multi-dimensional cell signal modeling. We suggest that the inherent compartmental nature of cells is a significant contributor to the splitting error in modeling cells.

Unlike atmospheric problems, where the chemistry is reasonably smooth spatially (it is a function of temperature, altitude, pressure, humidity, etc., all of which are continuous variables), cells contain clearly defined compartments and borders, over which chemistry can discontinuously vary. While more complex models of cell signaling may take into account such factors as temperature and pressure, which are indeed important for a multitude of cell functionality, the inclusion of locality into even the simplest models of cell signaling requires the presence of compartments and physical borders. Such structures create abrupt sources and sinks within the model chemistry, which substantially contribute to the splitting error.

As shown by Lanser and Verwer [225], a complete and general splitting error calculation for reaction-advection-diffusion problems is too cumbersome to be of much practical use. As such they prescribe a number of simplifications to make the calculation more manageable. Since we are not currently modeling advection, the splitting expression is simplified considerably. While this simplification makes the calculation more feasible, it is still at a significant time cost per step. In our simulations, we have only seen significant calculation time advantages for problems where uniform distributions of reactants reduce the role played by diffusion. In such situations the time step size can increase dramatically (as seen in the reaction error-dominated and diffusion error-dominated examples), and significantly reduce the calculation time. For more common heterogeneous problems, calculating the various truncation errors every N steps reduces the calculation cost of each by a factor of N, although this makes the method's time adaptivity less robust. The time cost to calculate reaction and diffusion errors using the methods presented in this paper are near negligible.

Currently the most common approach for modeling intracellular signaling networks is based on purely kinetic reaction systems modeled using ODEs, such as Gepasi [246] and Genesis [247]. This is referred to as the "well mixed model" as it assumes homogenous chemistry throughout the cell. Focus has also recently extended to modeling transport phenomena requiring the use of spatially resolved stochastic approaches such as MCell [248] and StochSim [249] as well as spatially resolved kinetic approaches like Virtual Cell [250] and CellSim [8]. The last of these approaches, or possibly even a hybrid of the last two methods, is the most suited for utilization of the Strang splitting error monitoring method outlined in this paper.

Other adaptive methods for solving partial differential equations include adaptive mesh methods [251] and multigrid methods [252], time adaptive splitting methods for quickly equilibrating diffusion [253], and *in situ* adaptive tabulation [207], essentially a storage and retrieval system. The main benefit of our method is ease of implementation into software packages that use fixed Cartesian grids. The method is independent of the underlying reaction and diffusion integrators, and for already existing implementations of Strang splitting only the splitting error calculation and the simple adaptive control method need to be added. However, we have recently proposed methods to extend splitting error control to adaptive mesh refinement techniques, providing adaptivity in space as well as time.

As a possible extension to this work, there are important intracellular advective processes such as active transport along actin filaments, which are interesting in so far they can be entirely decoupled from diffusion [254]. In other words, reactants undergoing such advection are not simultaneously diffusing. Adding an advection propagator to the Strang split calculation for such reactants introduces splitting error due to the coupling of reaction and advection, while for such reactants diffusion no longer contributes to the splitting error. As such, determining the significance of advection-reaction splitting error in reaction-diffusion/advection cell models is one possible step for further investigation into the role of splitting error in spatiotemporal cell modeling.

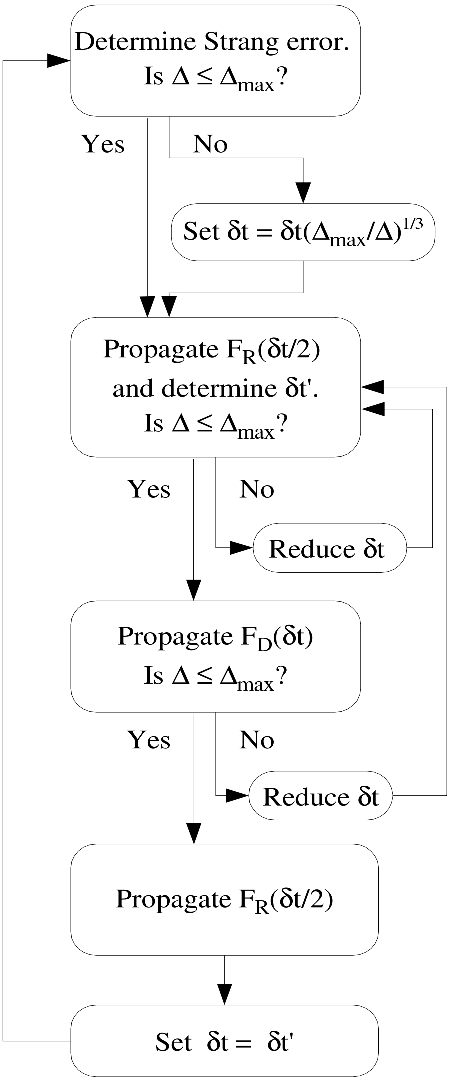


Figure 5.1: Algorithm flowchart for a single time step δt. The method consists of four steps: 1. Determine the Strang splitting error and adjust the time step, if needed. 2. Propagate the reaction operator a half time step. 3. Propagate the diffusion operator a full time step. 4. Propagate the reaction operator a half time step. The method checks to see if each error estimate is below the maximum tolerance set for its corresponding propagator. The time step choice δt is considered successful if all error estimates pass this check.

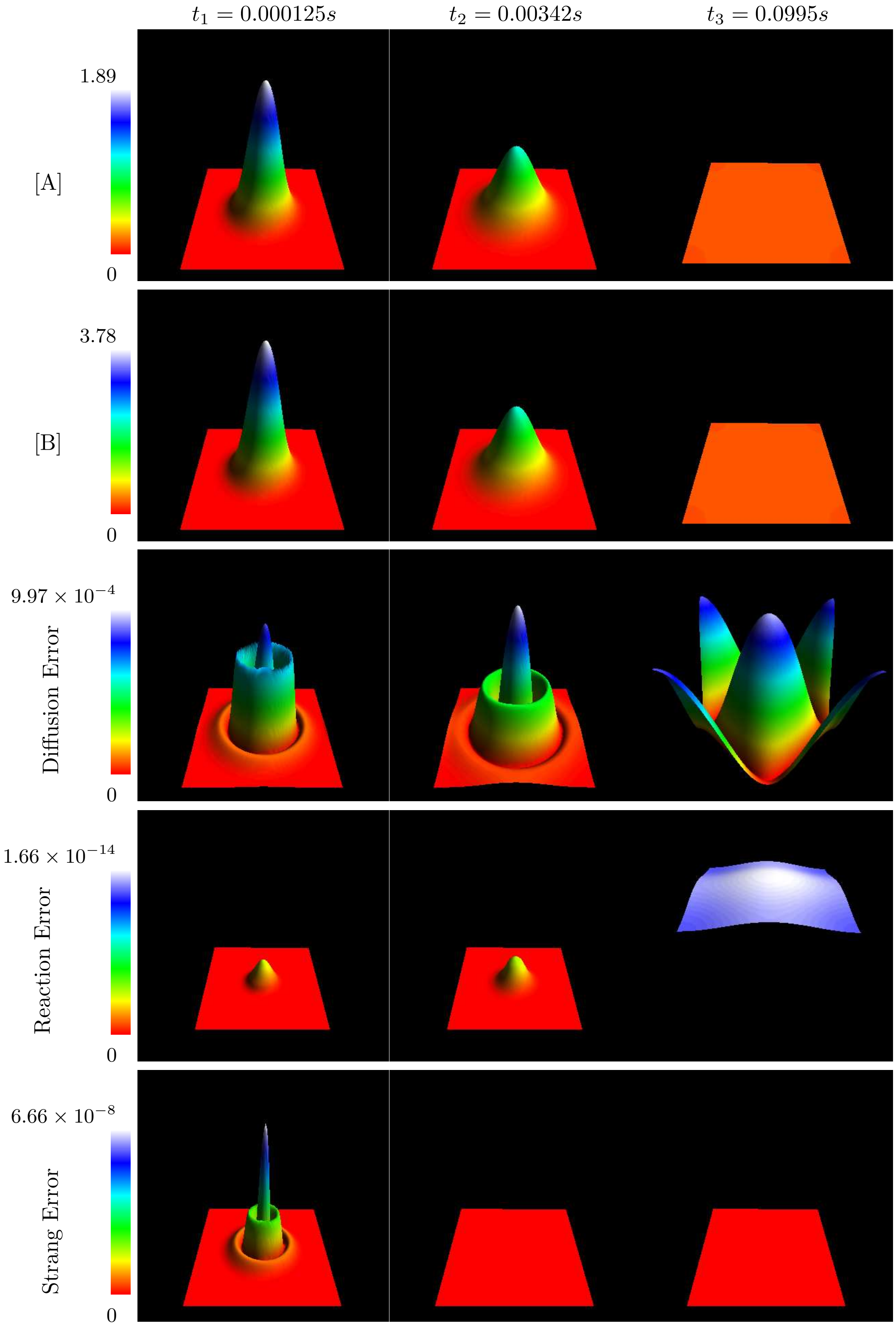


Figure 5.2: Detailed concentration and truncation error profiles of the diffusion error-dominated system during the initial 0.1 seconds of simulation. During this initial period, the Gaussian distributions of the reactants A and B quickly diffuse out to uniform distributions over the grid. The three columns of images correspond to the simulation times $t_1 = 0.000125s, t_2 = 0.00342s$ and $t_3 = 0.0995s$. The top two rows show the time evolution of the reactants A and B, and the bottom three rows show the pointwise maximum truncation error profiles due to diffusion, reaction, and Strang splitting.

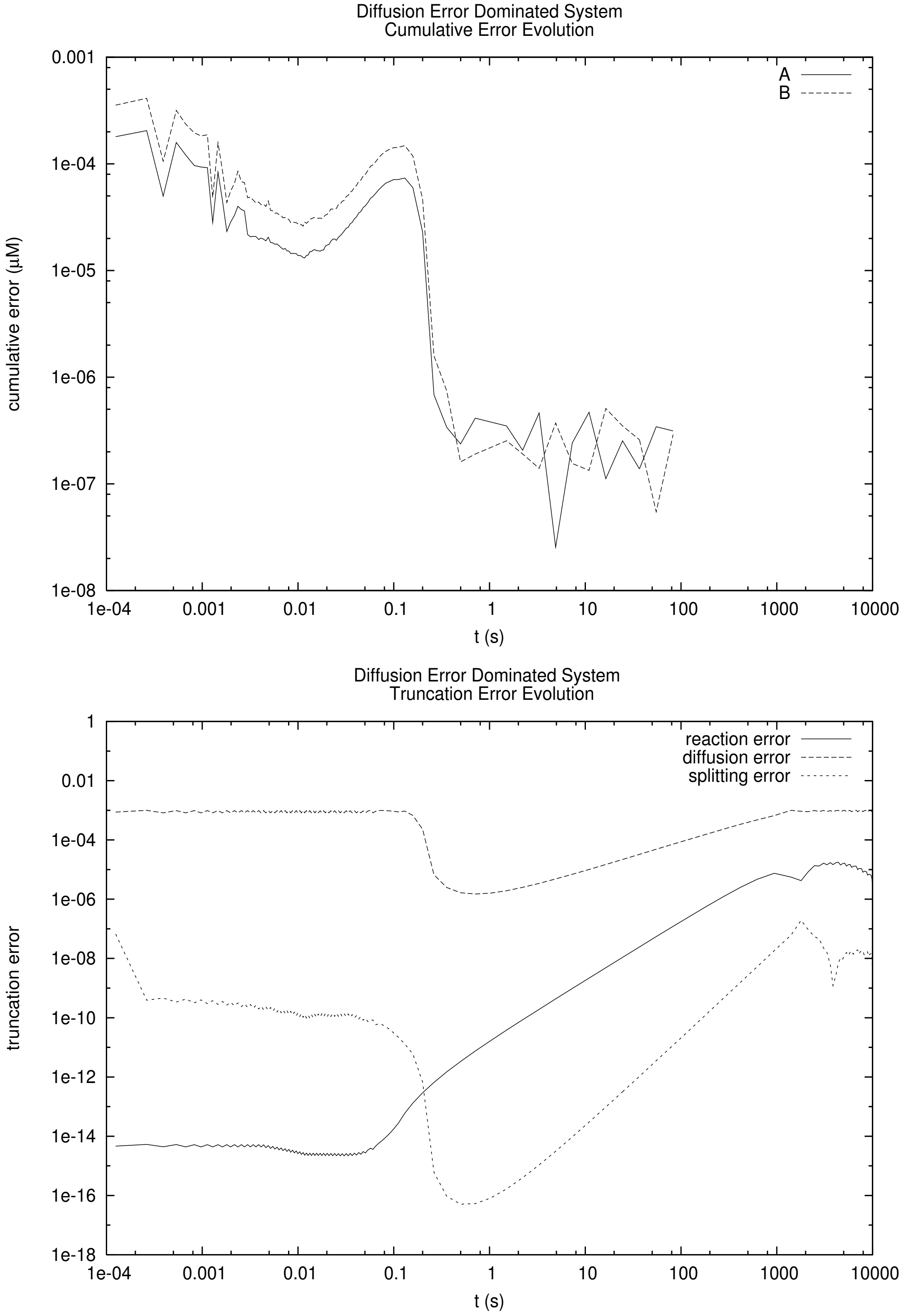


Figure 5.3: Maximal truncation error estimates (bottom) and cumulative errors (top) for the diffusion error-dominated system. After 0.1 seconds, the reactants A and B are uniformly distributed over the grid, resulting in reduced truncation error estimates for small time steps. The cumulative error also drops at this point, as the system reaches steady state.

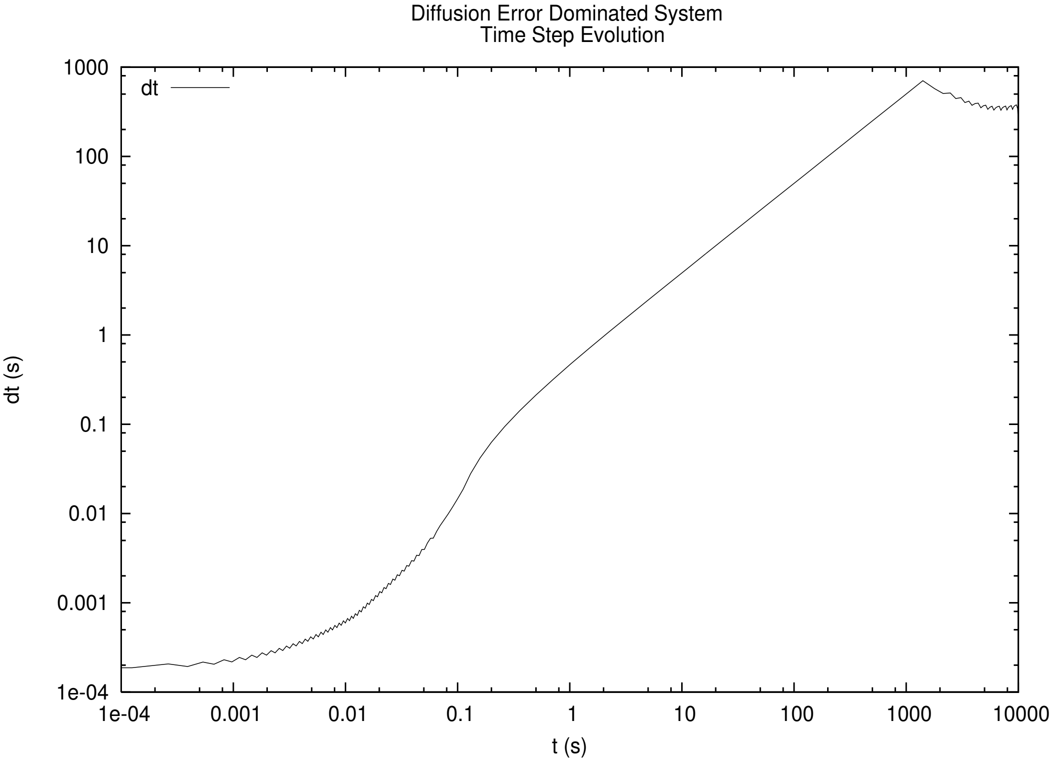


Figure 5.4: Evolution of the time step δt for the diffusion error-dominated system.

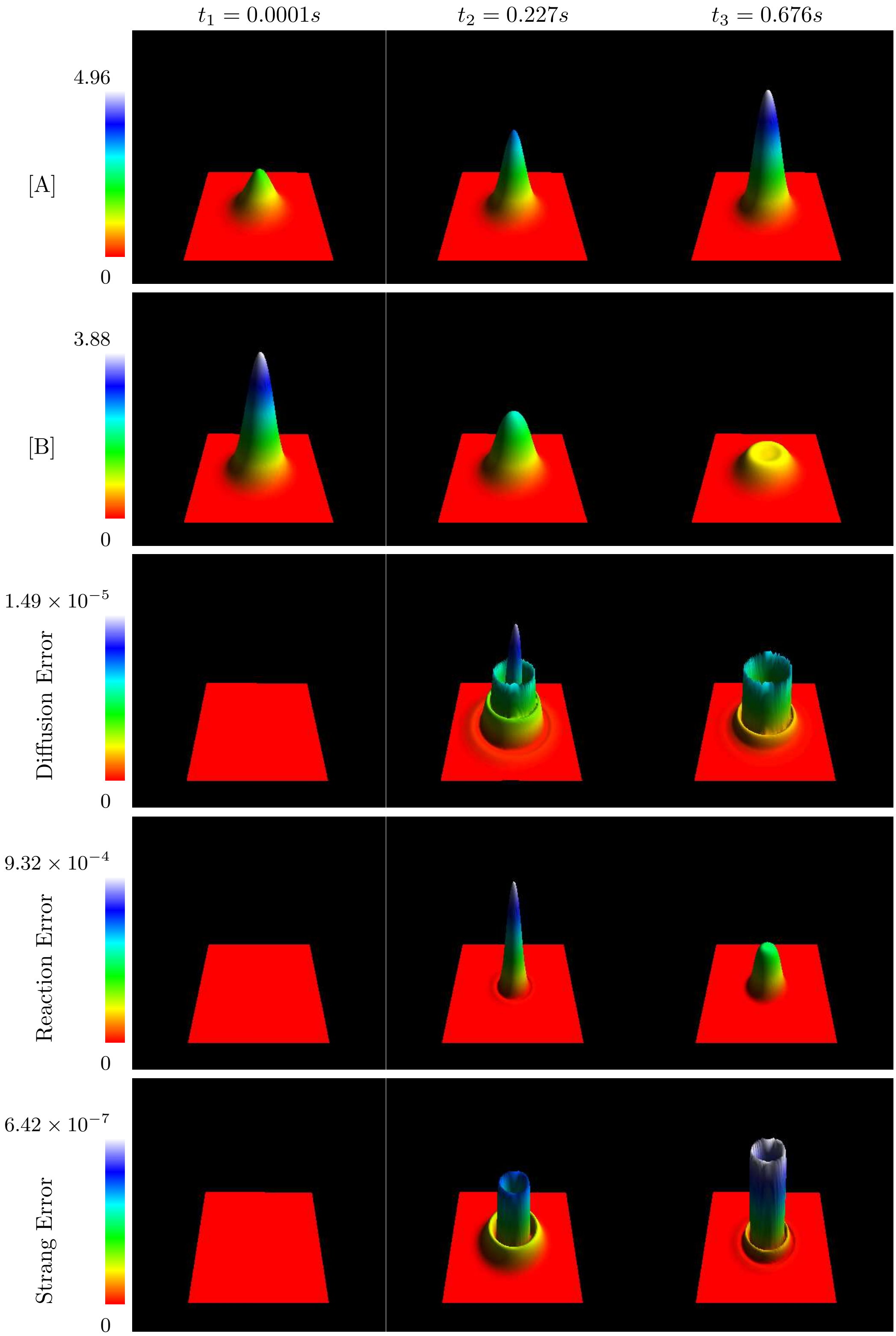


Figure 5.5: Detailed concentration and truncation error profiles of the reaction error-dominated system during the initial 0.676 seconds of simulation. During this initial period the Gaussian distributions of the reactants A and B react quickly to increase the concentration of A and reduce the concentration of B. The three columns of images correspond to the simulation times $t_1 = 0.0001s, t_2 = 0.227s$ and $t_3 = 0.676s$. The top two rows show the time evolution of the reactants A and B, and the bottom three rows show the maximum pointwise truncation error profiles due to diffusion, reaction, and Strang splitting.

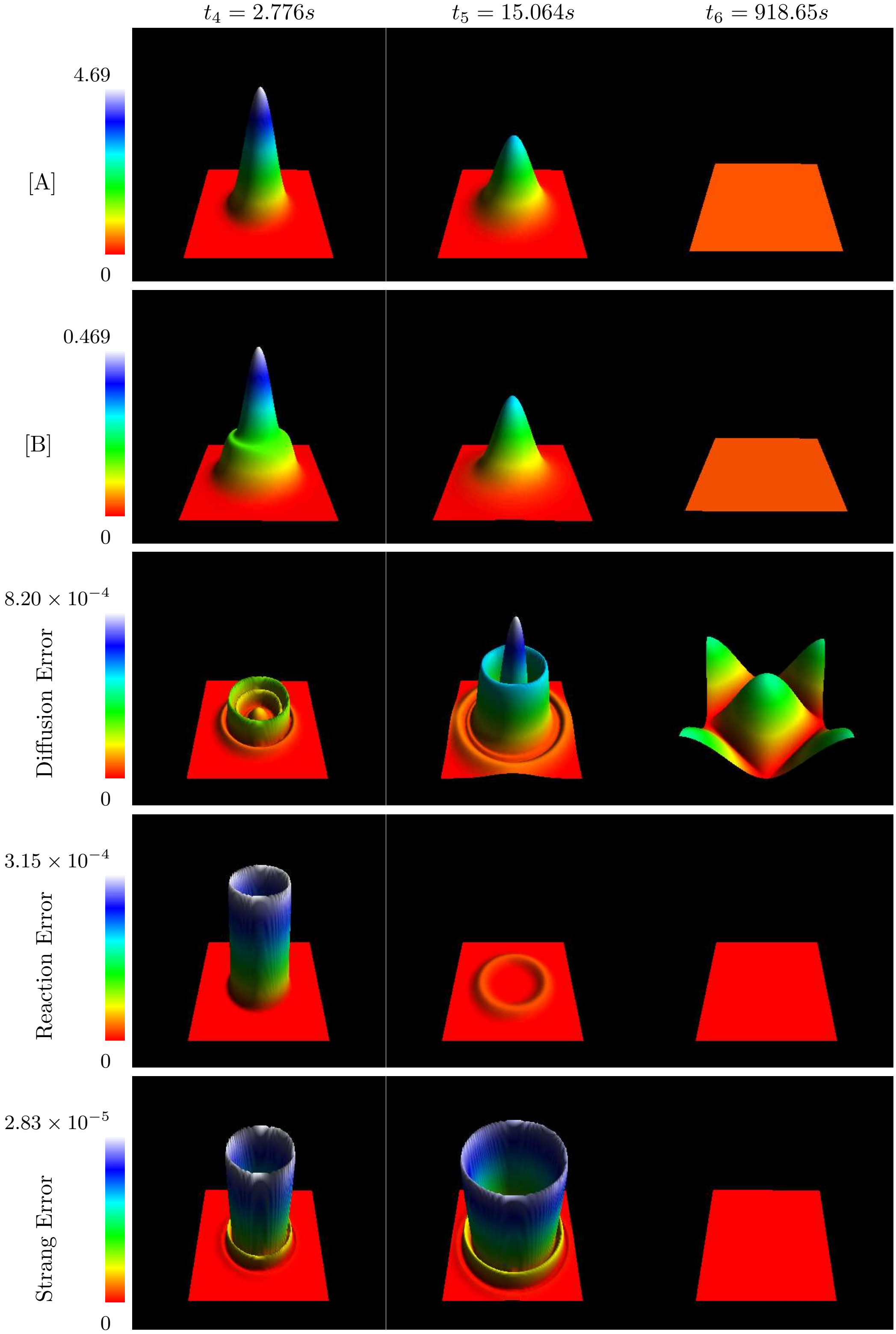


Figure 5.6: Detailed concentration and truncation error profiles of the reaction error-dominated system during the later, diffusion error-dominated, stages of simulation. During this period the reactants A and B slowly diffuse to a uniform distribution over the grid. The three columns of images correspond to the simulation times $t_4 = 2.776s, t_5 = 15.064s$ and $t_6 = 918.65s$. The top two rows show the time evolution of the reactants A and B, and the bottom three rows show the maximum pointwise truncation error profiles due to diffusion, reaction, and Strang splitting.

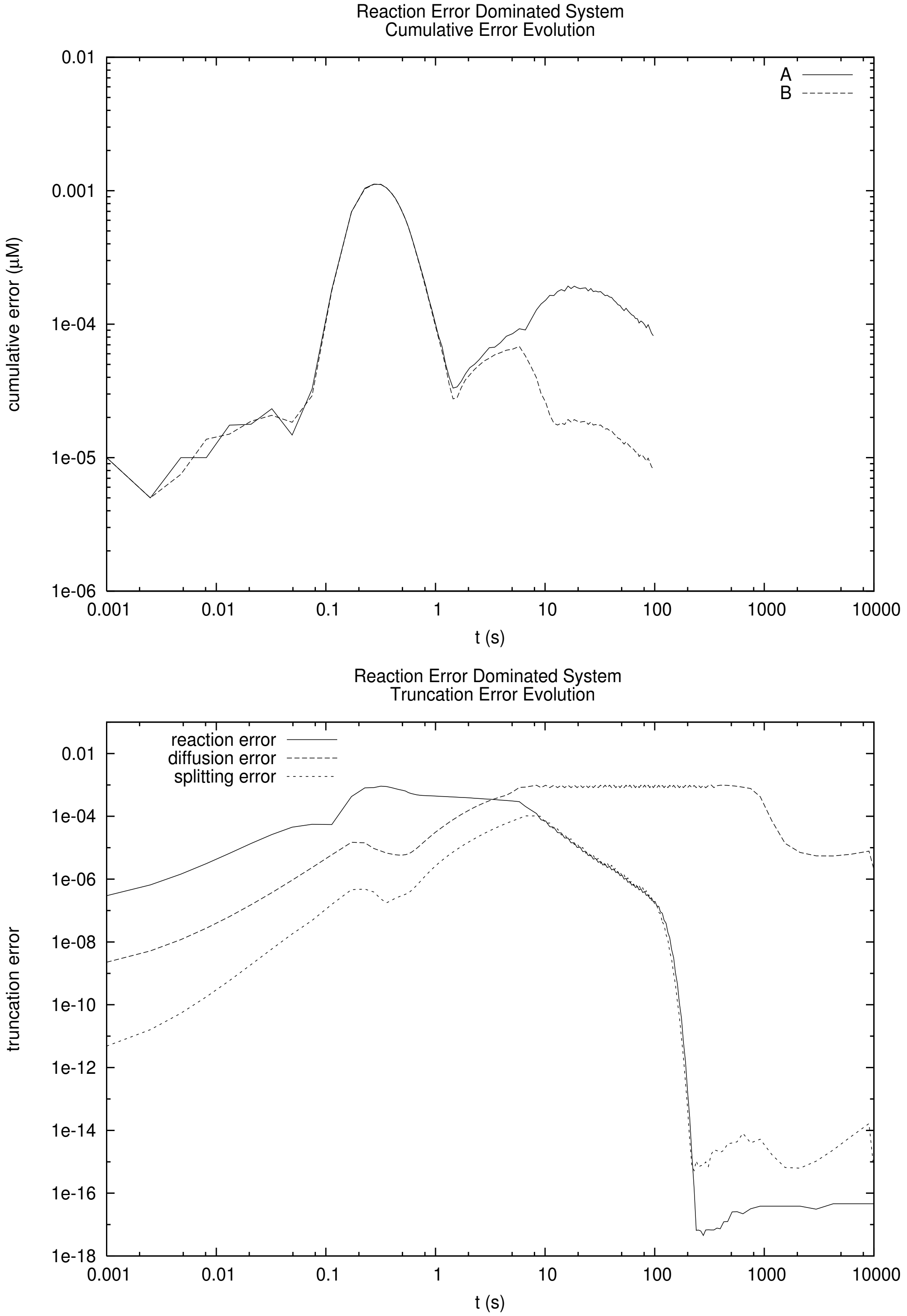


Figure 5.7: Truncation error estimates (bottom) and cumulative errors (top) in an initially reaction error-dominated system. The reactants A and B quickly reach kinetic reaction steady-state values at time $t \approx 15s$, by which point the system has become diffusion error-dominated. By time $t = 1000s$, the reactants have been uniformly distributed over the grid.

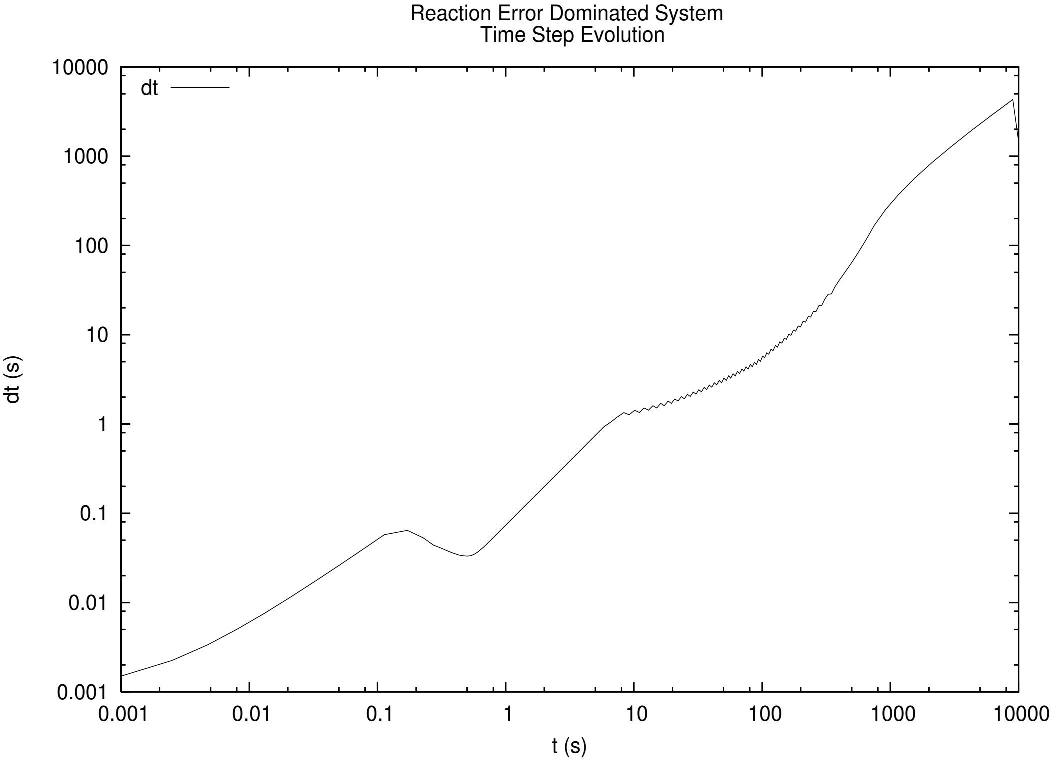


Figure 5.8: Evolution of the time step δt for the reaction error-dominated system.

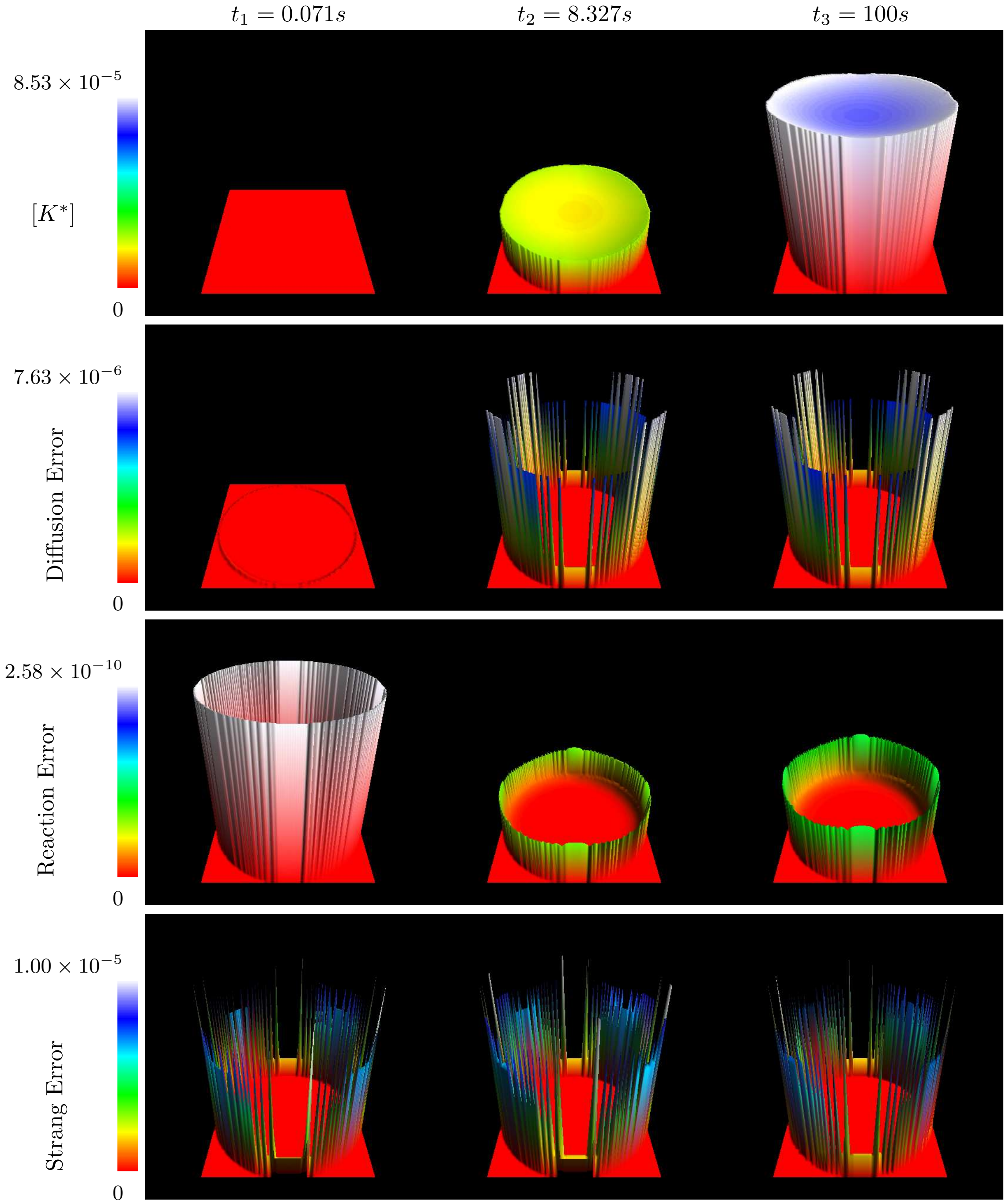
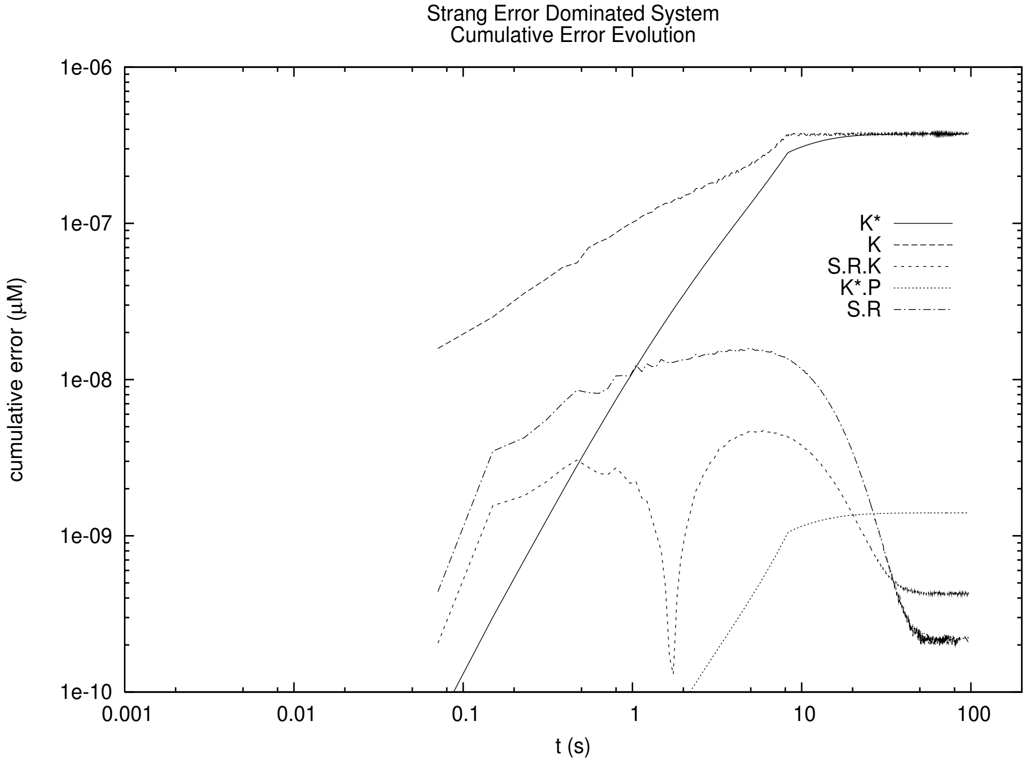


Figure 5.9: Activated kinase (K^*) concentration and truncation errors for the Strang splitting error-dominated system. The distribution of K^* rises to a steady state by time $t_3 = 100s$. The splitting error immediately increases during the initial time steps until it peaks at the maximum allowed value of $\Delta_{max} = 10^{-5}$. The splitting and diffusion errors are most prominent near the membrane region of the cell.



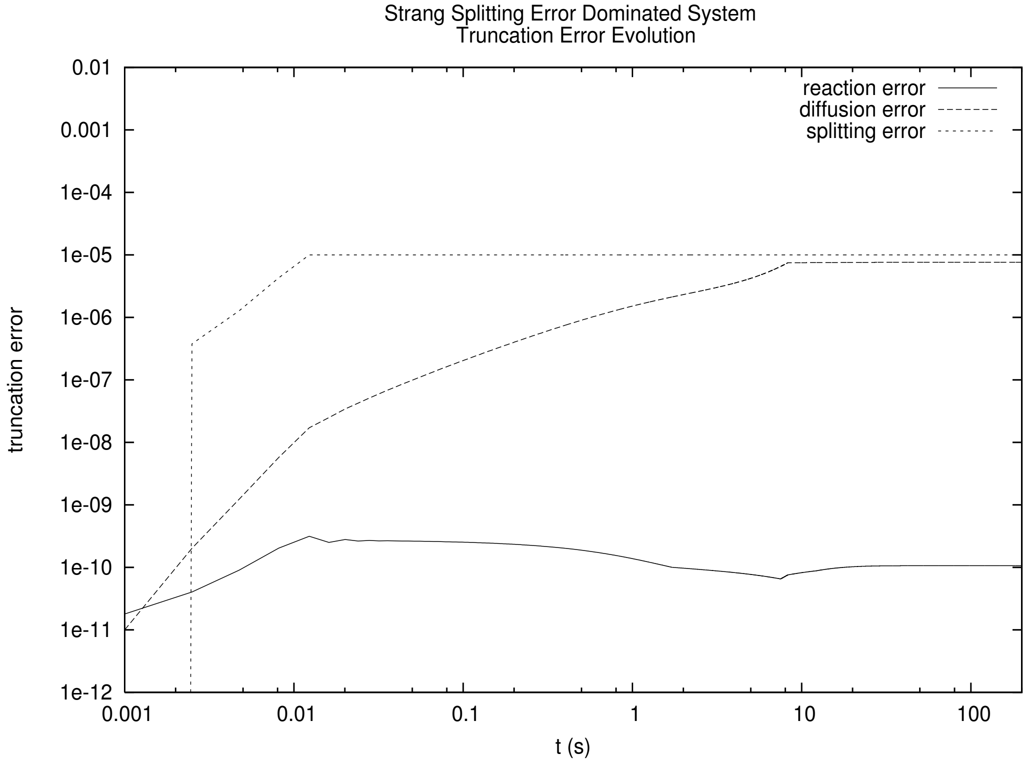


Figure 5.10: Truncation error estimates and cumulative errors of the Strang splitting error-dominated system. The diffusion error rapidly approaches the splitting error initially, then levels at 0.76 of the maximum allowed value.

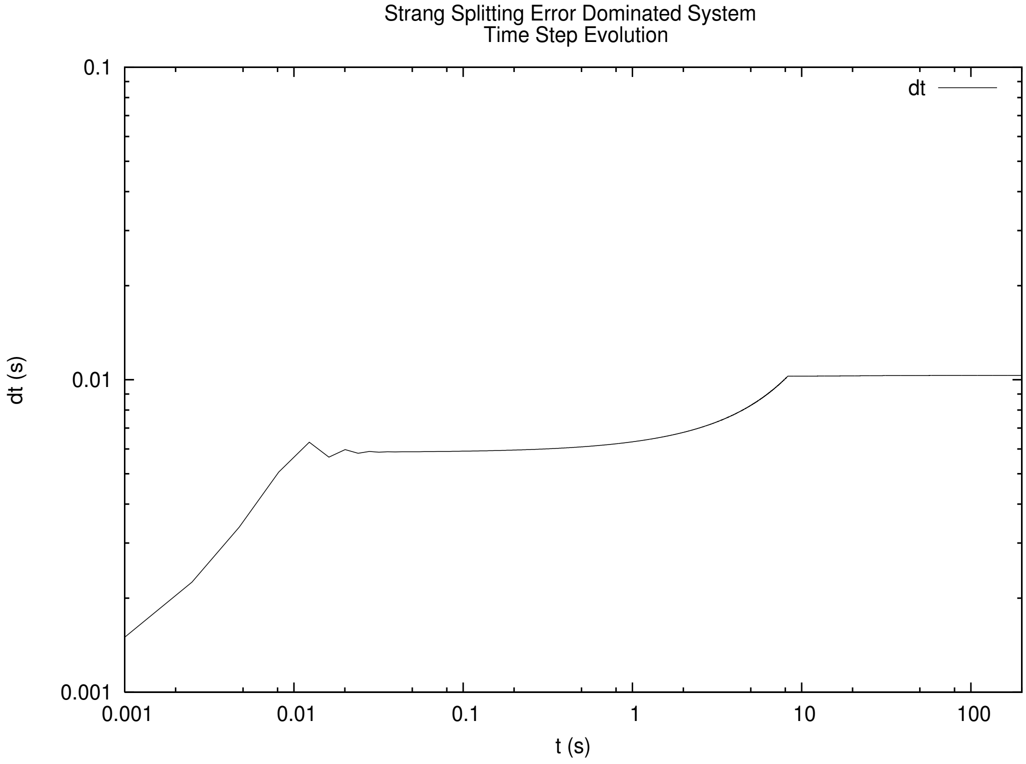


Figure 5.11: Evolution of the time step δt for the Strang splitting error-dominated system. The integration time step reaches a steady-state value of $0.01026s$ at time $t_2 = 8.327s$.

Bibliography

[1] Upinder S. Bhalla and Ravi Iyengar. Emergent properties of networks of biological signaling pathways. *Science*, 283:381–387, 1999.

[2] A. Ghosh, D. Pant, R. Zou, D. Miller, T. Shortell, H. Lin, and Y. Wang. `http://bio.physics.drexel.edu/research.html`, 2006. CellSim Home Page.

[3] A. Liwo, P. Arlukowicz, C. Czaplewski, S. Oldziej, J. Pillardy, and H.A. Scheraga. A method for optimizing potential-energy functions by hierarchical design of the potential-energy landscape: application to the unres force field. *PNAS*, 99:1937–42, 2002.

[4] Y. Komeiji, H. Yokoyama, M. Uebayasi, M. Taiji, T. Fukushige, D. Sugimoto, R. Takata, A. Shimizu, and K. Itsukashi. A high performance system for molecular dynamics simulation of biomolecules using a special-purpose computer. *Pac Symp. Biocomput.*, pages 472–87, 1996.

[5] R. Zou and A. Ghosh. Automated sensitivity analysis of stiff biochemical systems using a fourth-order adaptive step size rosenbrock integration method. *IEE Proc. Systems Biology*, 153:79–90, 2006.

[6] D.K. Pant and A. Ghosh. A systems biology approach for the study of cumulative oncogenes, with applications to the mapk signal transduction pathway. *Biophys. Chem.*, 119:49–60, 2006.

[7] D K Pant and A Ghosh. Automated oncogene detection in complex protein networks, with applications to the mapk signal transduction pathway. *Biophys. Chem.*, 113:275–288, 2005.

[8] Avijit Ghosh, David Miller, Rui Zou, Bahrad Sokhansanj, and A Kriete. *Computational Systems Biology*, chapter Spatio-Temporal Systems Biology, pages 327–362. Academic Press, Burlington, 2006.

[9] Avijit Ghosh, David Miller, Rui Zou, Bahrad Sokhansanj, and A Kriete. Integrated spatio-temporal model of cell signaling. Santa Barbara, 2005. Foundations of Systems Biology in Engineering.

[10] T. Joneson and D. Bar-Sagi. Ras effectors and their role in mitogenesis and oncogenesis. *J. Mol. Med.*, 75:587–93, 1997.

[11] C.J. Marshall. Ras effectors. *Curr. Opin. Cell. Biol.*, 8:197–204, 1996.

[12] A.F. Chambers and A.B. Tuck. Ras-responsive genes and tumor metastasis. *Crit Rev Oncog*, 4:95–114, 1993.

[13] A B Vojtek and C J Der. Increasing complexity of the ras signaling pathway. *J. Biol. Chem.*, 273:19925–19928, 1998.

[14] D.T. Denhardt. Signal-transducing protein phosphorylation cascades mediated by ras/rho proteins in the mammalian cell: the potential for multiplex signalling. *Biochem J*, 318:729–47, 1996.

[15] B.M. Marte and J. Downward. Pkb/akt: connecting phosphoinositide 3-kinase to cell survival and beyond. *Trends Biochem Sci*, 22:355–8, 1997.

[16] T. Tsuchida, H. Kijima, S. Hori, Y. Oshika, T. Tokunaga, K. Kawai, H. Yamazaki, Y. Ueyama, K.J. Scanlon, N. Tamaoki, and M. Nakamura. Adenovirus-mediated anti-k-ras ribozyme induces apoptosis and growth suppression of human pancreatic carcinoma. *Cancer Gene Ther*, 7:373–83, 2000.

[17] F. Uberall, K. Hellbert, S. Kampfer, K. Maly, A. Villunger, M. Spitaler, J. Mwanjewe, G. Baier-Bitterlich, G. Baier, and H. H. Grunicke. Evidence that atypical protein kinase c-lambda and atypical protein kinase c-zeta participate in ras-mediated reorganization of the f-actin cytoskeleton. *J. Cell. Biol.*, 144:413–425, 1999.

[18] S. Kampfer, F. Uberall, S. Giselbrecht, K. Hellbert, G. Baier, and H. H. Grunicke. Characterization of pkc isozyme specific functions in cellular signaling. *Adv Enzyme Regul.*, 38:35–48, 1998.

[19] S. J. Mansour, W. T. Matten, A. S. Hermann, J. M. Candia, S. Rong, K. Fukasawa, G. F. Vande Woude, and N. G. Ahn. Transformation of mammalian cells by constitutively active map kinase kinase. *Science*, 265:966–70, 1994.

[20] L. F. Parada, C. J. Tabin, C. Shih, and R. A. Weinberg. Human ej bladder carcinoma oncogene is homologue of harvey sarcoma virus ras gene. *Nature*, 297:474–8, 1982.

[21] C. J. Tabin, S. M. Bradley, C. I. Bargmann, R. A. Weinberg, A. G. Papageorge, E. M. Scolnick, R. Dhar, D. R. Lowy, and E. H. Chang. Mechanism of activation of a human oncogene. *Nature*, 300:143–9, 1982.

[22] R. W. Ellis, D. DeFeo, J. M. Maryak, H. A. Young, T. Y. Shih, E. H. Chang, D. R. Lowy, and E. M. Scolnick. Dual evolutionary origin for the rat genetic sequences of harvey murine sarcoma virus. *J Virol*, 36:408–20, 1980.

[23] J L Bos. Ras oncogenes in human cancer: a review. *Cancer Res.*, 49:4682–4689, 1989.

[24] R.A. Weinberg. Oncogenes, antioncogenes, and the molecular bases of multistep carcinogenesis. *Cancer Res*, 49:3713–21, 1989.

[25] R.G. Frank. New estimates of drug development costs. *J. Health Econ.*, 22:325–30, 2003.

[26] Upinder S. Bhalla, Prahlad T. Ram, and Ravi Iyengar. Map kinase phosphatase as a locus of flexibility in a mitogen-activated protein kinase signaling network. *Science*, 297:1018–1023, 2002.

[27] K. Kondoh, S. Torii, and E. Nishida. Control of map kinase signaling to the nucleus. *Chromosoma*, 114:86–91, 2005.

[28] J. Alberola-Ila and G. Hernandez-Hoyos. The ras/mapk cascade and the control of positive selection. *Immunol Rev*, 191:79–96, 2003.

[29] S. Boldt and W. Kolch. Targeting mapk signalling: Prometheus' fire or pandora's box? *Curr Pharm Des*, 10:1885–905, 2004.

[30] S. Boldt, U. H. Weidle, and W. Kolch. The role of mapk pathways in the action of chemotherapeutic drugs. *Carcinogenesis*, 23:1831–8, 2002.

[31] Gezhi Weng, Upinder S. Bhalla, and Ravi Iyengar. Complexity in biological signaling systems. *Science*, 284:92–96, 1999.

[32] J. D. Jordan and R. Iyengar. Modes of interactions between signaling pathways. *Biochem Pharmacol*, 55:1347–52, 1998.

[33] K. Giehl, B. Skripczynski, A. Mansard, A. Menke, and P. Gierschik. Growth factor-dependent activation of the ras-raf-mek-mapk pathway in the human pancreatic carcinoma cell line panc-1 carrying activated k-ras: implications for cell proliferation and cell migration. *Oncogene*, 19:2930–42, 2000.

[34] J. M. Olson and A. R. Hallahan. p38 map kinase: a convergence point in cancer therapy. *Trends Mol Med*, 10:125–9, 2004.

[35] Y. Suzuki, M. Orita, M. Shiraishi, K. Hayashi, and T. Sekiya. Detection of ras gene mutations in human lung cancers by single-strand conformation polymorphism analysis of polymerase chain reaction products. *Oncogene*, 5:1037–43, 1990.

[36] J. J. Oudejans, R. J. Slebos, F. A. Zoetmulder, W. J. Mooi, and S. Rodenhuis. Differential activation of ras genes by point mutation in human colon cancer with metastases to either lung or liver. *Int J Cancer*, 49:875–9, 1991.

[37] R. J. Slebos, J. A. Hoppin, P. E. Tolbert, E. A. Holly, J. W. Brock, R. H. Zhang, P. M. Bracci, J. Foley, P. Stockton, L. M. McGregor, G. P. Flake, and J. A. Taylor. K-ras and p53 in pancreatic cancer: association with medical history, histopathology, and environmental exposures in a population-based study. *Cancer Epidemiol Biomarkers Prev*, 9:1223–32, 2000.

[38] G. Gasparini, W. J. Gullick, P. Bevilacqua, J. R. Sainsbury, S. Meli, P. Boracchi, A. Testolin, G. La Malfa, and F. Pozza. Human breast cancer: prognostic significance of the c-erbb-2 oncoprotein compared with epidermal growth factor receptor, dna ploidy, and conventional pathologic features. *J Clin Oncol*, 10:686–95, 1992.

[39] Birgit Schoeberl, Claudia Eichler-Jonsson, Ernst Dieter Gilles, and Gretraud Muller. Computational modeling of the dynamics of the map kinase cascade activated by surface and internalized egf receptors. *Nature Biotechnology*, 20:370–375, 2002.

[40] B. N. Kholodenko. Negative feedback and ultrasensivity can bring about oscillations in the mitogen-activated protein kinase cascades. *Eur. J. Biochem.*, 267:1583–1488, 2000.

[41] A. Levchenko, J. Bruck, and P.W. Sternberg. Scaffold proteins may biphasically affect the levels of mitogen-activated protein kinase signaling and reduce its threshold properties. *PNAS*, 97:5818–5823, 2000.

[42] N. Thompson and J. Lyons. Recent progress in targeting the raf/mek/erk pathway with inhibitors in cancer drug discovery. *Curr Opin Pharmacol*, 5:350–6, 2005.

[43] N. M. Appels, J. H. Beijnen, and J. H. Schellens. Development of farnesyl transferase inhibitors: a review. *Oncologist*, 10:565–78, 2005.

[44] J. S. Macdonald, S. McCoy, R. P. Whitehead, S. Iqbal, J. L. Wade III, J. K. Giguere, and J. L. Abbruzzese. A phase ii study of farnesyl transferase inhibitor r115777 in pancreatic cancer: a southwest oncology group (swog 9924) study. *Invest New Drugs*, 23:485–7, 2005.

[45] H. Ren, S. K. Tai, F. Khuri, Z. Chu, and L. Mao. Farnesyltransferase inhibitor sch66336 induces rapid phosphorylation of eukaryotic translation elongation factor 2 in head and neck squamous cell carcinoma cells. *Cancer Res*, 65:5841–7, 2005.

[46] M. Beeram, A. Patnaik, and E. K. Rowinsky. Raf: a strategic target for therapeutic development against cancer. *J Clin Oncol*, 23:6771–90, 2005.

[47] S. S. Sridhar, D. Hedley, and L. L. Siu. Raf kinase as a target for anticancer therapeutics. *Mol Cancer Ther*, 4:677–85, 2005.

[48] E. M. Wallace, J. P. Lyssikatos, T. Yeh, J. D. Winkler, and K. Koch. Progress towards therapeutic small molecule mek inhibitors for use in cancer therapy. *Curr Top Med Chem*, 5:215–29, 2005.

[49] H. Ogiso, R. Ishitani, O. Nureki, S. Fukai, M. Yamanaka, J. H. Kim, K. Saito, A. Sakamoto, M. Inoue, M. Shirouzu, and S. Yokoyama. Crystal structure of the complex of human epidermal growth factor and receptor extracellular domains. *Cell*, 110:775–87, 2002.

[50] P. P. Roux and J. Blenis. Erk and p38 mapk-activated protein kinases: a family of protein kinases with diverse biological functions. *Microbiol Mol Biol Rev*, 68:320–44, 2004.

[51] G. Scita, P. Tenca, E. Frittoli, A. Tocchetti, M. Innocenti, G. Giardina, and P. P. Di Fiore. Signaling from ras to rac and beyond: not just a matter of gefs. *Embo J*, 19:2393–8, 2000.

[52] U. S. Bhalla and R. Iyengar. Robustness of the bistable behavior of a biological signaling feedback loop. *Chaos*, 11:221–226, 2001.

[53] M. K. Dougherty, J. Muller, D. A. Ritt, M. Zhou, X. Z. Zhou, T. D. Copeland, T. P. Conrads, T. D. Veenstra, K. P. Lu, and D. K. Morrison. Regulation of raf-1 by direct feedback phosphorylation. *Mol Cell*, 17:215–24, 2005.

[54] A. B. Vojtek and C. J. Der. Increasing complexity of the ras signaling pathway. *J Biol Chem*, 273:19925–8, 1998.

[55] W. Kolch. Meaningful relationships: the regulation of the ras/raf/mek/erk pathway by protein interactions. *Biochem J*, 351 Pt 2:289–305, 2000.

[56] L. Yanez, J. Groffen, and D. M. Valenzuela. c-k-ras mutations in human carcinomas occur preferentially in codon 12. *Oncogene*, 1:315–8, 1987.

[57] M. Homma and Y. Homma. [effect of phospholipase c inhibitor peptides on cancer cell growth]. *Gan To Kagaku Ryoho*, 24:156–63, 1997.

[58] M. K. Homma, Y. Homma, M. Yamasaki, S. Ohmi-Imajoh, and Y. Yuasa. Growth inhibition by phospholipase c inhibitor peptides of colorectal carcinoma cells derived from familial adenomatous polyposis. *Cell Growth Differ*, 7:281–8, 1996.

[59] A. L. Evdonin, I. V. Guzhova, B. A. Margulis, and N. D. Medvedeva. Phospholipse c inhibitor, u73122, stimulates release of hsp-70 stress protein from a431 human carcinoma cells. *Cancer Cell Int*, 4:2, 2004.

[60] J. Kassis, J. Moellinger, H. Lo, N. M. Greenberg, H. G. Kim, and A. Wells. A role for phospholipase c-gamma-mediated signaling in tumor cell invasion. *Clin Cancer Res*, 5:2251–60, 1999.

[61] W. K. Oh, H. Oh, B. Y. Kim, B. S. Kim, and J. S. Ahn. Crm-51006, a new phospholipase c (plc) inhibitor, produced by unidentified fungal strain mt51005. *J Antibiot (Tokyo)*, 57:808–11, 2004.

[62] H. Mogami, C. Lloyd Mills, and D. V. Gallacher. Phospholipase c inhibitor, u73122, releases intracellular ca2+, potentiates ins(1,4,5)p3-mediated ca2+ release and directly activates ion channels in mouse pancreatic acinar cells. *Biochem J*, 324 (Pt 2):645–51, 1997.

[63] B. R. Kelemen, K. Hsiao, and S. A. Goueli. Selective in vivo inhibition of mitogen-activated protein kinase activation using cell-permeable peptides. *J Biol Chem*, 277:8741–8, 2002.

[64] M. Knockaert, P. Lenormand, N. Gray, P. Schultz, J. Pouyssegur, and L. Meijer. p42/p44 mapks are intracellular targets of the cdk inhibitor purvalanol. *Oncogene*, 21:6413–24, 2002.

[65] J. Koivunen, V. Aaltonen, S. Koskela, P. Lehenkari, M. Laato, and J. Peltonen. Protein kinase c alpha/beta inhibitor go6976 promotes formation of cell junctions and inhibits invasion of urinary bladder carcinoma cells. *Cancer Res*, 64:5693–701, 2004.

[66] G. K. Schwartz, J. Jiang, D. Kelsen, and A. P. Albino. Protein kinase c: a novel target for inhibiting gastric cancer cell invasion. *J Natl Cancer Inst*, 85:402–7, 1993.

[67] C. A. Carter. Protein kinase c as a drug target: implications for drug or diet prevention and treatment of cancer. *Curr Drug Targets*, 1:163–83, 2000.

[68] M. Serova, A. Ghoul, K. A. Benhadji, E. Cvitkovic, S. Faivre, F. Calvo, F. Lokiec, and E. Raymond. Preclinical and clinical development of novel agents that target the protein kinase c family. *Semin Oncol*, 33:466–78, 2006.

[69] P. A. Philip and A. L. Harris. Potential for protein kinase c inhibitors in cancer therapy. *Cancer Treat Res*, 78:3–27, 1995.

[70] K. Horgan, E. Cooke, M. B. Hallett, and R. E. Mansel. Inhibition of protein kinase c mediated signal transduction by tamoxifen. importance for antitumour activity. *Biochem Pharmacol*, 35:4463–5, 1986.

[71] G. H. Baltuch, W. T. Couldwell, J. G. Villemure, and V. W. Yong. Protein kinase c inhibitors suppress cell growth in established and low-passage glioma cell lines. a comparison between staurosporine and tamoxifen. *Neurosurgery*, 33:495–501; discussion 501, 1993.

[72] P. de Medina, G. Favre, and M. Poirot. Multiple targeting by the antitumor drug tamoxifen: a structure-activity study. *Curr Med Chem Anticancer Agents*, 4:491–508, 2004.

[73] D. M. Springer, B. Y. Luh, J. J. Bronson, K. E. McElhone, M. M. Mansuri, K. R. Gregor, D. O. Nettleton, P. L. Stanley, and K. M. Tramposch. Biaryl diacid inhibitors of human s-pla2 with anti-inflammatory activity. *Bioorg Med Chem*, 8:1087–109, 2000.

[74] K. B. Glaser, R. P. Carlson, A. Sung, J. Bauer, Y. W. Lock, D. Holloway, R. Sturm, D. Hartman, T. Walter, S. Woeppel, and et al. Pharmacological characterization of way-121,520: a potent anti-inflammatory indomethacin-based inhibitor of 5-lipoxygenase (5-lo)/phospholipase a2 (pla2). *Agents Actions*, 39 Spec No:C30–2, 1993.

[75] J. P. Laye and J. H. Gill. Phospholipase a2 expression in tumours: a target for therapeutic intervention? *Drug Discov Today*, 8:710–6, 2003.

[76] P. L. de Souza, M. Castillo, and C. E. Myers. Enhancement of paclitaxel activity against hormone-refractory prostate cancer cells in vitro and in vivo by quinacrine. *Br J Cancer*, 75:1593–600, 1997.

[77] D. C. Sokal, A. Dabancens, R. Guzman-Serani, and J. Zipper. Cancer risk among women sterilized with transcervical quinacrine in chile: an update through 1996. *Fertil Steril*, 74:169–71, 2000.

[78] D. C. Sokal, J. Zipper, R. Guzman-Serani, and T. E. Aldrich. Cancer risk among women sterilized with transcervical quinacrine hydrochloride pellets, 1977 to 1991. *Fertil Steril*, 64:325–34, 1995.

[79] A. Wittinghofer and C. Herrmann. Ras-effector interactions, the problem of specificity. *FEBS Lett*, 369:52–6, 1995.

[80] J. S. Sebolt-Leopold. Mek inhibitors: a therapeutic approach to targeting the ras-map kinase pathway in tumors. *Curr Pharm Des*, 10:1907–14, 2004.

[81] J. S. Sebolt-Leopold. Development of anticancer drugs targeting the map kinase pathway. *Oncogene*, 19:6594–9, 2000.

[82] T. Suthiphongchai, S. Phimsen, U. Sakulkhu, and R. Tohtong. Pd98059-inhibited invasion of dunning rat prostate cancer cells involves suppression of motility but not mmp-2 or upa secretion. *Oncol Rep*, 15:1605–10, 2006.

[83] S. Zelivianski, M. Spellman, M. Kellerman, V. Kakitelashvilli, X. W. Zhou, E. Lugo, M. S. Lee, R. Taylor, T. L. Davis, R. Hauke, and M. F. Lin. Erk inhibitor pd98059 enhances docetaxel-induced apoptosis of androgen-independent human prostate cancer cells. *Int J Cancer*, 107:478–85, 2003.

[84] F. Caponigro, M. Casale, and J. Bryce. Farnesyl transferase inhibitors in clinical development. *Expert Opin Investig Drugs*, 12:943–54, 2003.

[85] S. R. Johnston. Farnesyl transferase inhibitors: a novel targeted tnerapy for cancer. *Lancet Oncol*, 2:18–26, 2001.

[86] S. R. Johnston and L. R. Kelland. Farnesyl transferase inhibitors–a novel therapy for breast cancer. *Endocr Relat Cancer*, 8:227–35, 2001.

[87] R. Yao, Y. Wang, Y. Lu, W. J. Lemon, D. W. End, C. J. Grubbs, R. A. Lubet, and M. You. Efficacy of the farnesyltransferase inhibitor r115777 in a rat mammary tumor model: role of haras mutations and use of microarray analysis in identifying potential targets. *Carcinogenesis*, 27:1420–31, 2006.

[88] J. J. Wright, K. Zerivitz, A. E. Gravell, and B. D. Cheson. Clinical trials referral resource. current clinical trials of r115777 (zarnestra). *Oncology (Williston Park)*, 16:930–1, 935–7, 2002.

[89] R. P. Whitehead, S. McCoy, J. S. Macdonald, S. E. Rivkin, M. A. Neubauer, S. R. Dakhil, H. J. Lenz, M. S. Tanaka, and J. L. Abbruzzese. Phase ii trial of r115777 (nsc #70818) in patients with advanced colorectal cancer: a southwest oncology group study. *Invest New Drugs*, 24:335–41, 2006.

[90] F. Warnberg, D. White, E. Anderson, F. Knox, R. B. Clarke, J. Morris, and N. J. Bundred. Effect of a farnesyl transferase inhibitor (r115777) on ductal carcinoma in situ of the breast in a human xenograft model and on breast and ovarian cancer cell growth in vitro and in vivo. *Breast Cancer Res*, 8:R21, 2006.

[91] W. S. Siegel-Lakhai, M. Crul, S. Zhang, R. W. Sparidans, D. Pluim, A. Howes, B. Solanki, J. H. Beijnen, and J. H. Schellens. Phase i and pharmacological study of the farnesyltransferase inhibitor tipifarnib (zarnestra, r115777) in combination with gemcitabine and cisplatin in patients with advanced solid tumours. *Br J Cancer*, 93:1222–9, 2005.

[92] R. A. Lubet, K. Christov, M. You, R. Yao, V. E. Steele, D. W. End, M. M. Juliana, and C. J. Grubbs. Effects of the farnesyl transferase inhibitor r115777 (zarnestra) on mammary carcinogenesis: prevention, therapy, and role of haras mutations. *Mol Cancer Ther*, 5:1073–8, 2006.

[93] S. J. Cohen, L. Ho, S. Ranganathan, J. L. Abbruzzese, R. K. Alpaugh, M. Beard, N. L. Lewis, S. McLaughlin, A. Rogatko, J. J. Perez-Ruixo, A. M. Thistle, T. Verhaeghe, H. Wang, L. M. Weiner, J. J. Wright, G. R. Hudes, and N. J. Meropol. Phase ii and pharmacodynamic study of the farnesyltransferase inhibitor r115777 as initial therapy in patients with metastatic pancreatic adenocarcinoma. *J Clin Oncol*, 21:1301–6, 2003.

[94] J. Tabernero, F. Rojo, I. Marimon, M. Voi, J. Albanell, M. Guix, F. Vazquez, J. Carulla, M. Cooper, J. Andreu, A. Van Vreckem, J. Bellmunt, V. Manne, J. A. Manning, C. Garrido, E. Felip, J. M. Del Campo, M. Garcia, S. Valverde, and J. Baselga. Phase i pharmacokinetic and pharmacodynamic study of weekly 1-hour and 24-hour infusion bms-214662, a farnesyltransferase inhibitor, in patients with advanced solid tumors. *J Clin Oncol*, 23:2521–33, 2005.

[95] W. C. Rose, F. Y. Lee, C. R. Fairchild, M. Lynch, T. Monticello, R. A. Kramer, and V. Manne. Preclinical antitumor activity of bms-214662, a highly apoptotic and novel farnesyltransferase inhibitor. *Cancer Res*, 61:7507–17, 2001.

[96] V. Papadimitrakopoulou, S. Agelaki, H. T. Tran, M. Kies, R. Gagel, R. Zinner, E. Kim, G. Ayers, J. Wright, and F. Khuri. Phase i study of the farnesyltransferase inhibitor bms-214662 given weekly in patients with solid tumors. *Clin Cancer Res*, 11:4151–9, 2005.

[97] H. J. Mackay, R. Hoekstra, F. A. Eskens, W. J. Loos, D. Crawford, M. Voi, A. Van Vreckem, T. R. Evans, and J. Verweij. A phase i pharmacokinetic and pharmacodynamic study of the farnesyl transferase inhibitor bms-214662 in combination with cisplatin in patients with advanced solid tumors. *Clin Cancer Res*, 10:2636–44, 2004.

[98] S. R. Johnston. Bms-214662 (bristol-myers squibb). *IDrugs*, 6:72–8, 2003.

[99] J. P. Eder Jr., D. P. Ryan, L. Appleman, A. X. Zhu, T. Puchalski, X. He, D. S. Sonnichsen, M. Cooper, J. Wright, J. W. Clark, and J. G. Supko. Phase i clinical trial of the farnesyltransferase inhibitor bms-214662 administered as a weekly 24 h continuous intravenous infusion in patients with advanced solid tumors. *Cancer Chemother Pharmacol*, 58:107–16, 2006.

[100] G. K. Dy, L. M. Bruzek, G. A. Croghan, S. Mandrekar, C. Erlichman, P. Peethambaram, H. C. Pitot, L. J. Hanson, J. M. Reid, A. Furth, S. Cheng, R. E. Martell, S. H. Kaufmann, and A. A. Adjei. A phase i trial of the novel farnesyl protein transferase inhibitor, bms-214662, in combination with paclitaxel and carboplatin in patients with advanced cancer. *Clin Cancer Res*, 11:1877–83, 2005.

[101] J. Cortes, S. Faderl, E. Estey, R. Kurzrock, D. Thomas, M. Beran, G. Garcia-Manero, A. Ferrajoli, F. Giles, C. Koller, S. O'Brien, J. Wright, S. A. Bai, and H. Kantarjian. Phase i study of bms-214662, a farnesyl transferase inhibitor in patients with acute leukemias and high-risk myelodysplastic syndromes. *J Clin Oncol*, 23:2805–12, 2005.

[102] H. R. Ashar, L. James, K. Gray, D. Carr, M. McGuirk, E. Maxwell, S. Black, L. Armstrong, R. J. Doll, A. G. Taveras, W. R. Bishop, and P. Kirschmeier. The farnesyl transferase inhibitor sch 66336 induces a g(2) –¿ m or g(1) pause in sensitive human tumor cell lines. *Exp Cell Res*, 262:17–27, 2001.

[103] A. Awada, F. A. Eskens, M. Piccart, D. L. Cutler, A. van der Gaast, H. Bleiberg, J. Wanders, M. N. Faber, P. Statkevich, P. Fumoleau, and J. Verweij. Phase i and pharmacological study of the oral farnesyltransferase inhibitor sch 66336 given once daily to patients with advanced solid tumours. *Eur J Cancer*, 38:2272–8, 2002.

[104] F. A. Eskens, A. Awada, D. L. Cutler, M. J. de Jonge, G. P. Luyten, M. N. Faber, P. Statkevich, A. Sparreboom, J. Verweij, A. R. Hanauske, and M. Piccart. Phase i and pharmacokinetic study of the oral farnesyl transferase inhibitor sch 66336 given twice daily to patients with advanced solid tumors. *J Clin Oncol*, 19:1167–75, 2001.

[105] M. Liu, M. S. Bryant, J. Chen, S. Lee, B. Yaremko, P. Lipari, M. Malkowski, E. Ferrari, L. Nielsen, N. Prioli, J. Dell, D. Sinha, J. Syed, W. A. Korfmacher, A. A. Nomeir, C. C. Lin, L. Wang, A. G. Taveras, R. J. Doll, F. G. Njoroge, A. K. Mallams, S. Remiszewski, J. J. Catino, V. M. Girijavallabhan, W. R. Bishop, and et al. Antitumor activity of sch 66336, an orally bioavailable tricyclic inhibitor of farnesyl protein transferase, in human tumor xenograft models and wap-ras transgenic mice. *Cancer Res*, 58:4947–56, 1998.

[106] M. Loprevite, R. E. Favoni, A. De Cupis, T. Scolaro, C. Semino, P. Mazzanti, and A. Ardizzoni. In vitro study of farnesyltransferase inhibitor sch 66336, in combination with chemotherapy and radiation, in non-small cell lung cancer cell lines. *Oncol Rep*, 11:407–14, 2004.

[107] F. G. Njoroge, B. Vibulbhan, P. Pinto, C. L. Strickland, W. R. Bishop, P. Kirschmeir, V. Girijavallabhan, and A. K. Ganguly. Trihalobenzocycloheptapyridine analogues of sch 66336 as potent inhibitors of farnesyl protein transferase. *Bioorg Med Chem*, 11:139–43, 2003.

[108] T. Petit, E. Izbicka, R. A. Lawrence, W. R. Bishop, S. Weitman, and D. D. Von Hoff. Activity of sch 66336, a tricyclic farnesyltransferase inhibitor, against human tumor colony-forming units. *Ann Oncol*, 10:449–53, 1999.

[109] S. Sharma, N. Kemeny, D. P. Kelsen, D. Ilson, E. O'Reilly, S. Zaknoen, C. Baum, P. Statkevich, E. Hollywood, Y. Zhu, and L. B. Saltz. A phase ii trial of farnesyl protein transferase inhibitor sch 66336, given by twice-daily oral administration, in patients with metastatic colorectal cancer refractory to 5-fluorouracil and irinotecan. *Ann Oncol*, 13:1067–71, 2002.

[110] E. Winquist, M. J. Moore, K. N. Chi, D. S. Ernst, H. Hirte, S. North, J. Powers, W. Walsh, T. Boucher, R. Patton, and L. Seymour. A multinomial phase ii study of lonafarnib (sch 66336) in patients with refractory urothelial cancer. *Urol Oncol*, 23:143–9, 2005.

[111] J. J. Wright, K. Zerivitz, and A. Gravell. Clinical trials referral resource. current clinical trials of bay 43-9006, part 1. *Oncology (Williston Park)*, 19:499–502, 2005.

[112] S. Wilhelm and D. S. Chien. Bay 43-9006: preclinical data. *Curr Pharm Des*, 8:2255–7, 2002.

[113] F. K. Tong, S. Chow, and D. Hedley. Pharmacodynamic monitoring of bay 43-9006 (sorafenib) in phase i clinical trials involving solid tumor and aml/mds patients, using flow cytometry to monitor activation of the erk pathway in peripheral blood cells. *Cytometry B Clin Cytom*, 70:107–14, 2006.

[114] D. Strumberg, D. Voliotis, J. G. Moeller, R. A. Hilger, H. Richly, S. Kredtke, C. Beling, M. E. Scheulen, and S. Seeber. Results of phase i pharmacokinetic and pharmacodynamic studies of the raf kinase inhibitor bay 43-9006 in patients with solid tumors. *Int J Clin Pharmacol Ther*, 40:580–1, 2002.

[115] D. Strumberg, A. Awada, H. Hirte, J. W. Clark, S. Seeber, P. Piccart, E. Hofstra, D. Voliotis, O. Christensen, A. Brueckner, and B. Schwartz. Pooled safety analysis of bay 43-9006 (sorafenib) monotherapy in patients with advanced solid tumours: Is rash associated with treatment outcome? *Eur J Cancer*, 42:548–56, 2006.

[116] H. Richly, B. F. Henning, P. Kupsch, K. Passarge, M. Grubert, R. A. Hilger, O. Christensen, E. Brendel, B. Schwartz, M. Ludwig, C. Flashar, R. Voigtmann, M. E. Scheulen, S. Seeber, and D. Strumberg. Results of a phase i trial of sorafenib (bay 43-9006) in combination with doxorubicin in patients with refractory solid tumors. *Ann Oncol*, 17:866–73, 2006.

[117] Activity of the raf kinase inhibitor bay 43-9006 in patients with advanced solid tumors. *Clin Colorectal Cancer*, 3:16–8, 2003.

[118] A. W. Tolcher, L. Reyno, P. M. Venner, S. D. Ernst, M. Moore, R. S. Geary, K. Chi, S. Hall, W. Walsh, A. Dorr, and E. Eisenhauer. A randomized phase ii and pharmacokinetic study of the antisense oligonucleotides isis 3521 and isis 5132 in patients with hormone-refractory prostate cancer. *Clin Cancer Res*, 8:2530–5, 2002.

[119] J. P. Stevenson, K. S. Yao, M. Gallagher, D. Friedland, E. P. Mitchell, A. Cassella, B. Monia, T. J. Kwoh, R. Yu, J. Holmlund, F. A. Dorr, and P. J. O'Dwyer. Phase i clinical/pharmacokinetic and pharmacodynamic trial of the c-raf-1 antisense oligonucleotide isis 5132 (cgp 69846a). *J Clin Oncol*, 17:2227–36, 1999.

[120] C. M. Rudin, J. Holmlund, G. F. Fleming, S. Mani, W. M. Stadler, P. Schumm, B. P. Monia, J. F. Johnston, R. Geary, R. Z. Yu, T. J. Kwoh, F. A. Dorr, and M. J. Ratain. Phase i trial of isis 5132, an antisense oligonucleotide inhibitor of c-raf-1, administered by 24-hour weekly infusion to patients with advanced cancer. *Clin Cancer Res*, 7:1214–20, 2001.

[121] A. M. Oza, L. Elit, K. Swenerton, W. Faught, P. Ghatage, M. Carey, L. McIntosh, A. Dorr, J. T. Holmlund, and E. Eisenhauer. Phase ii study of cgp 69846a (isis 5132) in recurrent epithelial ovarian cancer: an ncic clinical trials group study (ncic ind.116). *Gynecol Oncol*, 89:129–33, 2003.

[122] P. J. O'Dwyer, J. P. Stevenson, M. Gallagher, A. Cassella, I. Vasilevskaya, B. P. Monia, J. Holmlund, F. A. Dorr, and K. S. Yao. c-raf-1 depletion and tumor responses in patients treated with the c-raf-1 antisense oligodeoxynucleotide isis 5132 (cgp 69846a). *Clin Cancer Res*, 5:3977–82, 1999.

[123] H. Danahay, J. Giddings, R. A. Christian, H. E. Moser, and J. A. Phillips. Distribution of a 20-mer phosphorothioate oligonucleotide, cgp69846a (isis 5132), into airway leukocytes and epithelial cells following intratracheal delivery to brown-norway rats. *Pharm Res*, 16:1542–9, 1999.

[124] C. C. Cunningham, J. T. Holmlund, J. H. Schiller, R. S. Geary, T. J. Kwoh, A. Dorr, and J. Nemunaitis. A phase i trial of c-raf kinase antisense oligonucleotide isis 5132 administered as a continuous intravenous infusion in patients with advanced cancer. *Clin Cancer Res*, 6:1626–31, 2000.

[125] M. C. Cripps, A. T. Figueredo, A. M. Oza, M. J. Taylor, A. L. Fields, J. T. Holmlund, L. W. McIntosh, R. S. Geary, and E. A. Eisenhauer. Phase ii randomized study of isis 3521 and isis 5132 in patients with locally advanced or metastatic colorectal cancer: a national cancer institute of canada clinical trials group study. *Clin Cancer Res*, 8:2188–92, 2002.

[126] B. Coudert, A. Anthoney, W. Fiedler, J. P. Droz, V. Dieras, M. Borner, J. F. Smyth, R. Morant, M. J. de Vries, M. Roelvink, and P. Fumoleau. Phase ii trial with isis 5132 in patients with small-cell (sclc) and non-small cell (nsclc) lung cancer. a european organization for research and treatment of cancer (eortc) early clinical studies group report. *Eur J Cancer*, 37:2194–8, 2001.

[127] M Zeghouf, B Guibert, J-C Zeeh, and J Cherfils. Arf, sec7 and brefeldin a: a model towards the therapeutic inhibition of guanine nucleotide exchange factors. *Biochemical Society Transactions*, 33:1265–1268, 2005.

[128] S Schmidt, S Diriong, J Méry, E Fabbrizio, and A Debant. Identification of the first rhogef inhibitor, tripalpha, which targets the rhoa-specific gef domain of trio. *FEBS Letters*, 523:35–42, 2002.

[129] J L Bos, H Rehmann, and A Wittinghofer. Gefs and gaps: Critical elements in the control of small g proteins. *Cell*, 129:865–877, 2007.

[130] M E Fernandez-Zapico, N C Gonzalez-Paz, E Weiss, D N Savoy, J R Molina, R Fonseca, T C Smyrk, S T Chari, R Urrutia, and D D Billadeau. Ectopic expression of vav1 reveals an unexpected role in pancreatic cancer tumorigenesis. *Cancer Cell*, 7:39–49, 2005.

[131] H Dell. Tumorigenesis: Growing through gefs. *Nature Reviews Cancer*, 5:156, 2005.

[132] H Boulaiz, J Prados, C Melguizo, Á M García, J A Marchal, J L Ramos, E Carrillo, C Vélez, and A Aranega. Inhibition of growth and induction of apoptosis in human breast cancer by transfection of gef gene. *British Journal of Cancer*, 2003:192–198, 89.

[133] M. Pahor. Cancer risk in users of calcium channel blockers. *Hypertension*, 30:1641; author reply 1642, 1997.

[134] M. Pahor. Calcium channel blockers, risk of cancer, and the challenge of assessing medication safety. *Tenn Med*, 90:367–9, 1997.

[135] J. R. Daling. Calcium channel blockers and cancer: is an association biologically plausible? *Am J Hypertens*, 9:713–4, 1996.

[136] R. Zimlichman and Y. Shoenfeld. [calcium blockers and cancer–is there any link between them?]. *Harefuah*, 132:876–8, 1997.

[137] Z. F. Zhang, R. C. Kurtz, G. P. Yu, M. Sun, and S. Harlap. Calcium-channel blockers and cancer. *Lancet*, 348:1166–7, 1996.

[138] I. Vermes, C. Haanen, C. Fox, and C. Reutelingsperger. Cancer risk in users of calcium channel blockers. *Hypertension*, 30:1641–2, 1997.

[139] P. Trenkwalder. Calcium-channel blockers and cancer. starnberg study on epidemiology of parkinsonism and hypertension in the elderly. *Lancet*, 348:1167–8, 1996.

[140] L. Rosenberg, R. S. Rao, J. R. Palmer, B. L. Strom, P. D. Stolley, A. G. Zauber, M. E. Warshauer, and S. Shapiro. Calcium channel blockers and the risk of cancer. *Jama*, 279:1000–4, 1998.

[141] J. H. Olsen, H. T. Sorensen, S. Friis, J. K. McLaughlin, F. H. Steffensen, G. L. Nielsen, M. Andersen, J. F. Fraumeni Jr., and J. Olsen. Cancer risk in users of calcium channel blockers. *Hypertension*, 29:1091–4, 1997.

[142] S. Oliver. Calcium-channel blockers and cancer. *Lancet*, 348:1165; author reply 1167, 1996.

[143] S. J. Mann. Calcium-channel blockers and cancer. *Lancet*, 348:1165; author reply 1167, 1996.

[144] S. B. Kritchevsky and M. Pahor. Calcium-channel blockers and risk of cancer. *Lancet*, 349:1400, 1997.

[145] H. Jick, S. Jick, L. E. Derby, C. Vasilakis, M. W. Myers, and C. R. Meier. Calcium-channel blockers and risk of cancer. *Lancet*, 349:525–8, 1997.

[146] H. Jick. Calcium-channel blockers and risk of cancer. *Lancet*, 349:1699–700, 1997.

[147] M. A. James. Calcium-channel blockers and cancer. *Lancet*, 348:1168, 1996.

[148] H. L. Elliott and P. A. Meredith. Calcium-channel blockers and cancer. *Lancet*, 348:1165–6; author reply 1167, 1996.

[149] M. Brown. Calcium-channel blockers and cancer. *Lancet*, 348:1166; author reply 1167, 1996.

[150] H. J. Ahr, E. Bomhard, H. Mager, and G. Schluter. Calcium channel blockers and cancer: is there preclinical evidence for an association? *Cardiology*, 88 Suppl 3:68–72, 1997.

[151] Calcium channel blockers and breast cancer. *Harv Heart Lett*, 8:7, 1998.

[152] G. Lindberg, U. Lindblad, B. Low-Larsen, J. Merlo, A. Melander, and L. Rastam. Use of calcium channel blockers as antihypertensives in relation to mortality and cancer incidence: a population-based observational study. *Pharmacoepidemiol Drug Saf*, 11:493–7, 2002.

[153] H. J. Cohen, C. F. Pieper, J. T. Hanlon, W. E. Wall, B. M. Burchett, and R. J. Havlik. Calcium channel blockers and cancer. *Am J Med*, 108:210–5, 2000.

[154] M. Stahl, C. J. Bulpitt, A. J. Palmer, D. G. Beevers, E. C. Coles, and J. Webster. Calcium channel blockers, ace inhibitors, and the risk of cancer in hypertensive patients: a report from the department of health hypertension care computing project (dhccp). *J Hum Hypertens*, 14:299–304, 2000.

[155] H. T. Sorensen, J. H. Olsen, L. Mellemkjaer, A. Marie, F. H. Steffensen, J. K. McLaughlin, and J. A. Baron. Cancer risk and mortality in users of calcium channel blockers. a cohort study. *Cancer*, 89:165–70, 2000.

[156] J. D. Debes, R. O. Roberts, D. J. Jacobson, C. J. Girman, M. M. Lieber, D. J. Tindall, and S. J. Jacobsen. Inverse association between prostate cancer and the use of calcium channel blockers. *Cancer Epidemiol Biomarkers Prev*, 13:255–9, 2004.

[157] J. Y. Lee, S. J. Park, S. J. Park, M. J. Lee, H. Rhim, S. H. Seo, and K. S. Kim. Growth inhibition of human cancer cells in vitro by t-type calcium channel blockers. *Bioorg Med Chem Lett*, 2006.

[158] K. Sato, J. Ishizuka, C. W. Cooper, D. H. Chung, T. Tsuchiya, T. Uchida, S. Rajaraman, C. M. Townsend Jr., and J. C. Thompson. Inhibitory effect of calcium channel blockers on growth of pancreatic cancer cells. *Pancreas*, 9:193–202, 1994.

[159] A. W. Tolcher, K. H. Cowan, D. Solomon, F. Ognibene, B. Goldspiel, R. Chang, M. H. Noone, A. M. Denicoff, C. S. Barnes, M. R. Gossard, P. A. Fetsch, S. L. Berg, F. M. Balis, D. J. Venzon, and J. A. O'Shaughnessy. Phase i crossover study of paclitaxel with r-verapamil in patients with metastatic breast cancer. *J Clin Oncol*, 14:1173–84, 1996.

[160] C. V. Timcheva and D. K. Todorov. Does verapamil help overcome multidrug resistance in tumor cell lines and cancer patients? *J Chemother*, 8:295–9, 1996.

[161] C. W. Taylor, W. S. Dalton, K. Mosley, R. T. Dorr, and S. E. Salmon. Combination chemotherapy with cyclophosphamide, vincristine, adriamycin, and dexamethasone (cvad) plus oral quinine and verapamil in patients with advanced breast cancer. *Breast Cancer Res Treat*, 42:7–14, 1997.

[162] I. B. Shchepotin, D. A. McRae, M. Shabahang, R. R. Buras, and S. R. Evans. Hyperthermia and verapamil inhibit the growth of human colon cancer xenografts in vivo through apoptosis. *Anticancer Res*, 17:2213–6, 1997.

[163] S. Sela, S. R. Husain, J. W. Pearson, D. L. Longo, and A. Rahman. Reversal of multidrug resistance in human colon cancer cells expressing the human mdr1 gene by liposomes in combination with monoclonal antibody or verapamil. *J Natl Cancer Inst*, 87:123–8, 1995.

[164] V. Rybalchenko, N. Prevarskaya, F. Van Coppenolle, G. Legrand, L. Lemonnier, X. Le Bourhis, and R. Skryma. Verapamil inhibits proliferation of lncap human prostate cancer cells influencing k+ channel gating. *Mol Pharmacol*, 59:1376–87, 2001.

[165] K. Nakatani, T. Ezaki, K. Sakamoto, T. Sugizaki, and N. Miyagi. [effects of combination therapy of carmofur and nicardipine of human gastric cancer in nude mice]. *Gan To Kagaku Ryoho*, 12:969–71, 1985.

[166] M. Kurooka, T. Hongyo, H. Nakajima, R. Baskar, L. Y. Li, K. Fukuda, K. Sutoh, M. Miyata, H. Matsuda, and T. Nomura. High incidence of esophageal cancer in esophageal achalasia by the oral administration of n-amyl-n-methylnitrosamine and its prevention by nicardipine hydrochloride in mice. *Cancer Lett*, 127:55–61, 1998.

[167] C. Arisawa, Y. Kageyama, S. Kawakami, and K. Kihara. Tnp-470 combined with nicardipine suppresses in vivo growth of pc-3, a human prostate cancer cell line. *Urol Oncol*, 7:229–34, 2002.

[168] D. Miller and A. Ghosh. A fully adaptive reaction-diffusion integration scheme with applications to systems biology. *Journal of Computational Physics*, 226:1509–1531, 2007.

[169] Jing Yuan Fang and Bruce C Richardson. The mapk signalling pathways and colorectal cancer. *The Lancet Oncology*, 6:322–327, 2005.

[170] A S Coutts and L C Murphy. Elevated mitogen-activated protein kinase activity in estrogen-nonresponsive human breast cancer cells. *Cancer Research*, 58:4071–4074, 1998.

[171] T R Devereux, W Holliday, C Anna, N Ress, J Roycroft, and R C Sills. Map kinase activation correlates with k-ras mutation and loss of heterozygosity on chromosome 6 in alveolar bronchiolar carcinomas from b6c3f1 mice exposed to vanadium pentoxide for 2 years. *Carcinogenesis*, 23:1737–1743, 2002.

[172] A S Dhillon, S Hagan, O Rath, and W Kolch. Map kinase signalling pathways in cancer. *Oncogene*, 26:3279–3290, 2007.

[173] A. Kumar, R. Zou, and A. Ghosh. A systems biology approach for lead drug target prediction with applications to the mapk signal transduction pathway. 2006.

[174] A. Ghosh, D. Pant, R. Zou, A. Kumar, and D. Miller. From simulation to therapy: A systems biology approach to oncogene detection. In *7th International Conference on Systems Biology*, Yokohama, 2006.

[175] B Alberts, A Johnson, J Lewis, M Raff, K Roberts, and P Walter. *Molecular Biology of the Cell*. Garland Science, fourth edition, 2002.

[176] S Sprang. Gefs: master regulators of g-protein activation. *Trends in Biochemical Sciences*, 26:266–267, 2001.

[177] Guy C Brown and Boris N Kholodenko. Spatial gradients of cellular phospho-proteins. *FEBS Letters*, 457:452–454, 1999.

[178] Boris N Kholodenko, G C Brown, and et al. Diffusion control of protein phosphorylation in signal transduction pathways. *Biochem J*, 350:901–907, 2000.

[179] Boris N Kholodenko. Four-dimensional organization of protein kinase signaling cascades: the role of diffusion, endocytosis and molecular motors. *J. Exp. Biol*, 206:2073–2082, 2003.

[180] M. A. Peletier, H. V. Westerhoff, and et al. Control of spatially heterogeneous and time-varying cellular reaction networks: a new summation law. *J Theor Biol*, 225:477–487, 2003.

[181] S Khurana, S Kreydiyyeh, A Aronzon, W A Hoogerwerf, S G Rhee, M Donowitz, and M E Cohen. Asymmetric signal transduction in polarized ileal na(+)-absorbing cells: carbachol activates brush-border but not basolateral-membrane pip2-plc and translocates plc-gamma 1 only to the brush border. *Biochem Journal*, 313:509–518, 1996.

[182] T. L. Holdaway-Clarke, J. A. Feijo, and et al. Pollen tube growth and the intracellular cytosolic calcium gradient oscillate in phase while extracellular calcium influx is delayed. *Plant Cell*, 9:1999–2010, 1997.

[183] Hubert Lam, Jean-Yves Matroule, and Christine Jacobs-Wagner. The asymmetric spatial distribution of bacterial signal transduction proteins coordinates cell cycle events. *Developmental Cell*, 5:149–159, 2003.

[184] T. Y. Belenkaya, C. Han, and et al. Drosophila dpp morphogen movement is independent of dynamin-mediated endocytosis but regulated by the glypican members of heparan sulfate proteoglycans. *Cell*, 119:231–244, 2004.

[185] R Brown. A brief account of microscopical observations, etc. personal labbook, 1827.

[186] G Borman, F Brosens, and E DeSchutter. *Modeling Molecular Diffusion*, page 189. MIT Press, 2004.

[187] M Arrio-Dupont, G Foucault, and et al. Translational diffusion of globular proteins in the cytoplasm of cultured muscle cells. *Biophys J*, 78:901–907, 2000.

[188] D Purich, editor. *Enzymatic Kinetics and Mechanism: Detection and Characterization of Enzyme Reaction Intermediates: Methods in Enzymology*. Academic Press, 2004.

[189] H. H. McAdams and A. Arkin. Stochastic mechanisms in gene expression. *Proc Natl Acad Sci*, 94:814–819, 1997.

[190] H. H. McAdams and A. Arkin. It's a noisy business! genetic regulation at the nanomolar scale. *Trends Genet*, 15:65–69, 1999.

[191] M. B. Elowitz and S. Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403:335–338, 2000.

[192] H. H. McAdams and A. Arkin. Stochastic mechanisms in gene expression. *Annu Rev Biophys Biomol Struct*, 27:199–224, 1998.

[193] D T Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J Comp Phys.*, 22:403–434, 1976.

[194] D T Gillespie. Exact stochastic simulation of coupled chemical reactions. *Genetics*, 81:2340–2361, 1977.

[195] A Arkin, J. Ross, and et al. Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected escherichia coli cells. *Genetics*, 149:1633–1648, 1998.

[196] M. A. Gibson and J. Bruck. Efficient exact stochastic simulation of chemical systems with many species and many channels. *Journal of Physical Chemistry A*, 104:1876–1889, 2000.

[197] J J Lukkien, J P L Segers, and et al. Efficient monte carlo methods for the simulation of catalytic surface reactions. *Physical Review E*, 58:2598–2610, 1998.

[198] P Jung and G Mayer-Kress. Noise controlled spiral growth in excitable media. *Chaos*, 5:458–462, 1995.

[199] P Jung and G Mayer-Kress. Spatiotemporal stochastic resonance in excitable media. *Physical Review Letters*, 74:2130–2133, 1995.

[200] J Dunkel, S Hilbert, and et al. Stochastic resonance in biological nonlinear evolution models. *Phys Rev E*, 69:056118, 2004.

[201] I Z Kiss, J. L. Hudson, and et al. Noise-aided synchronization of coupled chaotic electrochemical oscillators. *Phys Rev E*, 70:026210, 2004.

[202] B Doiron, B Lindner, and et al. Oscillatory activity in electrosensory neurons increases with the spatial correlation of the stochastic input stimulus. *Phys Rev Lett*, 93:048101, 2004.

[203] S. Coombes and et al. R. Hinch. Receptors, sparks and waves in a fire-diffuse-fire framework for calcium release. *Prog Biophys Mol Biol*, 85:197–216, 2004.

[204] T E Turner, S. Schnell, and et al. Stochastic approaches for modelling in vivo reactions. *Comput Biol Chem*, 28:165–178, 2004.

[205] G Strang. On the construction and comparison of difference schemes. *SIAM Journal on Numerical Analysis*, 5:506–517, 1968.

[206] B. Sportisse. An analysis of operator splitting techniques in the stiff case. *Journal of Computational Physics*, 161:140–168, 2000.

[207] M. A. Singer and S. B. Pope. Exploiting isat to solve the reaction-diffusion equation. *Combustion Theory and Modelling*, 8:361–384, 2004.

[208] J Crank and P Nicolson. A practical method for numerical evaluation of solutions of partial differential equations of the heat-conduction type. *Proceedings of the Cambridge Philosophical Society*, 43:50–67, 1947.

[209] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes in C.* Cambridge University Press, UK, second edition, 1992.

[210] J Douglas. Alternating direction methods for three space variables. *Numerische Mathematik*, 4:41–63, 1962.

[211] J. G. Verwer, E. J. Spee, J. G. Blom, and W. H. Hundsdorfer. A second order rosenbrock method applied to photochemical dispersion problems. *SIAM Journal on Scientific Computation*, 20:1456–1480, 1999.

[212] Free Software Foundation. `http://www.gnu.org/copyleft/gpl.html`, 1989, 1991. GNU Public Licence.

[213] P Pacheco. *Parallel Programming with MPI.* Morgan Kaufmann, 1996.

[214] E Selkov. On the mechanism of single-frequency self-oscillations in glycolysis. i. a simple kinetic model. *Eur. J. Biochem*, 4:79–86, 1968.

[215] P Gray and S K Scott. Autocatalytic reactions in the isothermal, continuous stirred tank reactor : Isolas and other forms of multistability. *Chem.Eng.Sci*, 38:29–43, 1983.

[216] John E Pearson. Complex patterns in a simple system. *Science*, 261:189–192, 1993.

[217] M J Cole, M Pirity, and et al. Shedding light on bioscience. symposium on optical imaging: Applications to biology and medicine. *EMBO Rep*, 4:838–843, 2003.

[218] R F Murphy. Automated interpretation of protein subcellular location patterns: implication for early caner detection and assessment. *Ann N Y Acad Sci*, 1020:124–131, 2004.

[219] C Conrad, H Erfle, and et al. Automatic identification of subcellular phenotypes on human cell arrays. *Genome Res*, 14:1130–1136, 2004.

[220] M A DeBernardi and G Brooker. Simultaneous flourescence ratio imaging of cyclic amp and calcium kinetics in single living cells. *Adv Second Messenger Phosphoprotein Res*, 32:195–213, 1998.

[221] A M Smith and S Nie. Chemical analysis and cellular imaging with quantum dots. *AnalystAug*, 129:672–677, 2004.

[222] J Lippincott-Schwartz, E Snapp, and A Kenworthy. Studiying protein dynamics in living cells. *Nat Reviews Molecular Cell Biology*, 2:444–456, 2001.

[223] E A Jares-Erijman and T M Jovin. Fret imaging. *Nat Biotech*, 21:1387–1396, 2003.

[224] M Offterdinger, V Georget, A Girod, and P I H Bastians. Imaging phosphorylation dynamics of the epidermal growth factor receptor. *J Biol Chem*, 279:36972–36981, 2004.

[225] D. Lanser and J. G. Verwer. Analysis of operator splitting for advection-diffusion-reaction problems from air pollution modelling. *Journal of Computational and Applied Mathematics*, 111:201–216, 1999.

[226] R I McLachlan. On the numerical integration of ordinary differential equations by symmetric composition methods. *SIAM J. Sci. Comput.*, 16:151–168, 1995.

[227] R I McLachlan and G R W Quispel. Splitting methods. *Acta Numerica*, 11:341–434, 2002.

[228] R. Gilmore. *Lie Groups, Lie Algebras, and Some of Their Applications.* Krieger Publishing Company, Malabar, 1974.

[229] J. M. Sanz-Serna. *The State of the Art in Numerical Analysis*, chapter Geometric Integration, pages 121–143. Clarendon Press, Oxford, 1997.

[230] J. M. Sanz-Serna and M. P. Calvo. *Numerical Hamiltonian Problems.* Chapman and Hall, London, 1994.

[231] R Kozlov, A Kvearno, and B Owren. The behavior of the local error in splitting methods applied to stiff problems. *J. Comp. Phys.*, 195:576–593, 2004.

[232] S Descombes and M Massot. Operator splitting for nonlinear reaction-diffusion systems with an entropic structure: singular perturbation and order reduction. *Numerische Mathematik*, 97:667–698, 2004.

[233] J. E. Campbell. On the theory of continuous transformation groups. *Proceedings of the London Mathematical Society*, 28:381–390, 1897.

[234] E. B. Dynkin. On the representation of the series $log(e^x e^y)$ for non-commutative x and y by commutators. *Mathematicheskii Sbornik*, 25:155–162, 1949.

[235] R. Gilmore. Baker-campbell-hausdorff formulas. *Journal of Mathematical Physics*, 15:2090–2091, 1974.

[236] J. Douglas and J. Gunn. A general formulation of alternating direction methods. *Numerische Mathematik*, 6:428–453, 1964.

[237] K James and W Riha. Convergence criteria for successive overrelaxation. *SIAM J. Num. Analysis*, 12:137–143, 1975.

[238] R Reuter. Solving (cyclic) tridiagonal systems. *ACM SIGAPL APL Quote Quad*, 18:6–12, 1998.

[239] R. D. Richtmyer and K. W. Morton. *Difference Methods for Initial-Value Problems.* Interscience Publishers, NY, second edition, 1967.

[240] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes in C++.* Cambridge University Press, UK, 2002.

[241] Q Sheng. Global error estimates for exponential splitting. *IMA Journal of Numerical Analysis*, 14:27–56, 1993.

[242] W Bao and Jie Shen. A fourth-order time-splitting laguarre-hermite pseudospectal method for bose-einstein condensates. *SIAM J. Sci. Comp.*, 26:2010–2028, 2005.

[243] S Chin. Quantum statistical calculations and symplectic corrector algorithms. *Physical Review E*, 69:046118, 2004.

[244] W Bao and Chunxiong Zheng. A time-splitting spectral method for three-wave interactions in media with competing quadratic and cubic nonlinearities. *Commun. comput. phys.*, 2:123–140, 2007.

[245] K H Karlsen, K-A Lie, J R Natvig, H F Nordhaug, and H K Dahle. Operator splitting methods for systems of convection-diffusion equations: nonlinear error mechanisms and correction strategies. *J. Comp. Phys.*, 173:636–663, 2001.

[246] P Mendes. Biochemistry by numbers: simulation of biochemical pathways with gepasi 3. *Trends Biochem. Sci.*, 22:361–363, 1997.

[247] J. Bower and D. Beeman. *The Book of GENESIS: Exploring realistic Neural Models with the General Neural Simulation System.* Springer-Verlag, New York, 1994.

[248] Thomas M. Bartol Joel R. Stiles. *Computational Neuroscience: Realistic Modeling for Experimentalists,*, chapter Monte Carlo Methods for Simulating Realistic Synaptic Microphysiology Using MCell, pages 87–127. CRC Press, 2001.

[249] Nicolas Le Novre and Thomas Simon Shimizu. Stochsim: modelling of stochastic biomolecular processes. *Bioinformatics*, 17:575–576, 2001.

[250] Leslie M. Loew and James C. Schaff. The virtual cell: a software environment for computational cell biology. *Trends in Biotechnology*, 19:401–406, 2001.

[251] P. A. Zegeling and H. P. Kok. Adaptive moving mesh computations for reaction–diffusion systems. *J. Comput. Appl. Math.*, 168(1-2):519–528, 2004.

[252] Ching-Shan Chou, Yong-Tao Zhang, Rui Zhao, and Qing Nie. Numerical methods for stiff reaction-diffusion systems. *Discrete and continuous dynamical systems - Series B*, 7:515–525, 2007.

[253] James C. Schaff, Boris M. Slepchenko, Yung-Sze Choi, John Wagner, Diana Resasco, and Leslie M. Loew. Analysis of nonlinear dynamics on arbitrary geometetries with the virtual cell. *Chaos*, 11:115–131, 2001.

[254] A.-T. Dinh, T. Theofanous, and S. Mitragotri. A model for intracellular trafficking of adenoviral vectors. *Biophysical Journal*, 89:1574–1588, 2005.

Vita

Address: 885 N 23rd Street
Philadelphia, PA 19130
Phone: (215) 913 1906
E-mail: `djm65@drexel.edu`

Education	**Ph.D.** Physics, Drexel University Concentration in Computational Biophysics/Systems Biology	2008
	M.S. Physics, Drexel University	2004
	B.A. Physics, The College of Wooster Concentration in Nonlinear Dynamics/Cellular Automata	2002
	University of Aberdeen, Scotland	2000

Teaching	TA, Contemporary Physics I,II,III	Drexel University
	TA, Physics I,II,III	Drexel University
	TA, Introductory Physics I,II,III	Drexel University
	TA, Physical Foundations of Engineering	Drexel University
	TA, Energy II – Thermodynamics for Engineers	Drexel University

Publications

1. D. Miller, G. Lukas, A. Ghosh, “Reducing Side Effects: A Systems Approach for the Prediction of Wild Type MAPK Pathway Response to Targeted Drugs,” Proceedings, 8th International Conference on Systems Biology, Long Beach (2007).

2. D. Miller, A. Ghosh, “A Fully Adaptive Reaction-Diffusion Integration Scheme with Applications to Systems Biology,” Journal of Computational Physics, 226, 1509-1531 (2007).

3. A. Ghosh, D. Pant, R. Zou, A. Kumar, D. Miller, “From Simulation to Therapy: A Systems Biology Approach to Oncogene Detection,” Proceedings, 7th International Conference on Systems Biology, Yokohama, Japan (2006).

4. A. Ghosh, D. Pant, A. Kumar, R. Zou, D. Miller, “From Simulation to Therapy: A Systems Biology Approach to Oncogene Detection,” Proceedings, NIC Workshop 2006: From Computational Biophysics to Systems Biology, Jülich, Germany (2006).

5. A. Ghosh, D. Miller, R. Zou, B. Sokhansanj, A. Kriete, “Spatiotemporal Systems Biology,” published in *Computational Systems Biology*, A. Kriete, R. Eils (eds.), Academic Press : Burlington (2006).

6. A. Ghosh, D. Miller, R. Zou, H. Pais, B. Sokhansanj, A. Kriete, “Integrated Spatio-temporal Model of Cell Signaling,” Proceedings, Foundations of Systems Biology and Engineering, Santa Barbara (2005).

7. J. Lindner, S. Hughes, D. Miller, B. Thomas, K. Wiesenfeld, “The Flux Creep Automaton,” International Journal of Bifurcation and Chaos, 14, 1155-1175 (2004).