ABSTRACT

SAWYER, MEGAN ELIZABETH. Compartmentalizing the Sunlight Vitamin: Physiologically Based Pharmacokinetic Modeling and Vitamin D. (Under the direction of Hien T Tran and Marina V Evans.)

Increasing levels of vitamin D deficiency are seen in modern society, and as a result, use of vitamin D supplements are increasing. A question raised, and not yet answered, is how much supplementation would be enough. One method of answering this is through the use of physiologically-based pharmacokinetic (PBPK) modeling techniques to predict the effects of supplementation on a general population. In this work, we develop a PBPK model of the vitamin D metabolic cascade. The background of the model follows the standard PBPK model structure; however, we make several modifications to more accurately represent the observed biology of the vitamin D cascade. The resulting model considers daily dosing both through diet and supplementation. Parameter estimates are constrained to available literature bounds and the resulting PBPK model fits three separate dose levels.

Within the calibration of the PBPK model, we perform sensitivity analysis to determine which parameters most influence and contribute to model output. As sensitivity may not constitute identifiability, we use several techniques to select a subset of parameters that are estimable and identifiable, and optimize for model fit to available data over these parameters. Several optimization algorithms are considered.

Implications of this work include: Predictive capability for physiological dosage levels for subpopulations, such as age or gender-restricted populations; complete set-up for the addition of sunlight exposure routes; and the capability for coupling the PBPK model with exposure-prediction models.

Compartmentalizing the Sunlight Vitamin: Physiologically Based Pharmacokinetic Modeling and Vitamin D

by
Megan Elizabeth Sawyer

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Applied Mathematics

Raleigh, North Carolina

2013

APPROVED BY:

Alun Lloyd	John Franke
Hien T Tran Co-chair of Advisory Committee	Marina V Evans Co-chair of Advisory Committee

DEDICATION

To my parents, who not only taught me the stars are within reach,

but helped build a ladder to touch them.

BIOGRAPHY

Megan Sawyer was born in Denver, CO on November 2, 1984. She received the Bachelor of Science in Mathematics Education from University of Colorado Denver in 2007, graduating magna cum laude with academic distinction. She then completed the Post-Baccalaureate Program in Mathematics at Smith College in 2008 before starting her career at North Carolina State University. Here, Megan received the Master of Science in Mathematics in 2010 and will receive the Doctor of Philosophy in Applied Mathematics. Megan will be an Assistant Professor of Mathematics at Southern New Hampshire University in Manchester, NH.

ACKNOWLEDGEMENTS

There are several groups of individuals I would like to thank for their resolute support over the years. First and foremost, I'd like to thank my family for their unwavering belief in me and my capacity to learn. By providing me with the physical and emotional support I needed to succeed in school, you helped make it possible. I love you all.

I'd like to thank my advisors, not only my current ones, Drs. Marina Evans and Hien Tran, but also my previous advisors for each challenging me in different ways and helping me to become a better researcher and teacher. I'd also like to single out Dr. Ilse Ipsen for making one of my first courses at NCSU demanding in all the right ways. You are an inspiration for my teaching.

My friends, thank you so much for your encouragement, your support, your love, your presence. Without you, I know I would have been one lonely person. Jess and Michael, I have no words to express how very much I love you. My voices of reason, Kegan and Angel, I will never forget you and what you have taught me. Thank you.

To my mentor, Dr. Ruth Haas: That year at Smith was one of the best of my career. From that first email with the "Ma'am" to our coffee/beer/hiking talks, you have always been über-supportive in everything I do. You are a major reason why I stayed in school.

To Kate and Lindsay, and all those who have received a special piece of my heart over the past five years, I cherish my time with you. You have taught me so much about the difference between surviving and thriving in graduate school, and by extension, life. I love you deeply, and always will.

TABLE OF CONTENTS

LIST OF TABLES vii

LIST OF FIGURES viii

Chapter 1: Introduction 1

Chapter 2: Biology and Kinetics of Vitamin D 4
- **2.1 The Biology** 5
- **2.2 Kinetics** 7
  - **2.2.1 Initial Activation** 10
  - **2.2.2 25-Hydroxylation** 13
  - **2.2.3 1-Hydroxylation** 16
  - **2.2.4 Deactivation of Vitamin D** 18
  - **2.2.5 Regulation of the Vitamin D Pathway** 19
- **2.3 Summary** 21

Chapter 3: Model Development and Basic Analysis 22
- **3.1 PBPK Models** 24
  - **3.1.1 Mathematical Description** 25
- **3.2 Vitamin D Model** 26
  - **3.2.1 General Model** 26
  - **3.2.2 Study Design** 30
  - **3.2.3 Physiological Parameter Selection** 31
  - **3.2.4 Partition Coefficients** 32
  - **3.2.5 Kinetic Parameters** 37
  - **3.2.6 Absorption and Excretion** 42
- **3.3 Mathematical Model** 43
- **3.4 Existence and Uniqueness** 47
  - **3.4.1 Basic Model** 48
  - **3.4.2 Dynamic Adipose Partition Coefficients** 52
  - **3.4.3 Transcriptional Regulation of Enzyme Concentrations** 55

Chapter 4: Model Calibration and Subset Selection 57
- **4.1 Identifiability and Estimability** 58
  - **4.1.1 (Structural) Identifiability** 60
  - **4.1.2 Estimability** 62
  - **4.1.3 Sensitivity-Based Identifiability Analysis** 63
- **4.2 Parameter Subset Selection** 70
  - **4.2.1 Parameter Identifiabilty** 70

4.2.2 Parameter Estimability . . . 73
4.2.3 Optimization Methods . . . 74
4.3 Parameter Confidence Intervals . . . 76
4.4 Parameter Estimation . . . 78

Chapter 5: Model Results . . . 80
5.1 Initial Model Structure . . . 80
5.2 Dynamic Adipose Partition Coefficients . . . 82
5.3 Final Model and Optimized Parameters . . . 84

Chapter 6: Conclusion and Future Directions . . . 87

REFERENCES . . . 89

APPENDICES . . . 99
Appendix A: Model Development Supplement . . . 100
A.1 Symbols and Abbreviations . . . 100
A.2 Prediction of Initial Conditions . . . 102
Appendix B: Approximations and Optimizations . . . 106
B.1 Calculation of Sensitivity Coefficients . . . 106
B.1.1 Finite Difference Method . . . 107
B.1.2 Automatic Differentiation . . . 109
B.1.3 Comparison of Calculation Techniques . . . 110
B.2 Optimization Techniques . . . 110
B.2.1 Constraint Transformations . . . 112
B.2.2 Direct Search Methods . . . 114
B.2.3 Gradient-Based Methods . . . 118
Appendix C: Previous Work . . . 122
C.1 REU 2011: Metabolism of Bromochloromethane . . . 122
C.2 REU 2012: Dermal Exposure of Lindane . . . 126

LIST OF TABLES

Table 2.1 Primary reasons for compound choices 10
Table 2.2 Percent free of each substrate and the affinity of each to DBP. [1]As calculated using methods described in Feldman (1972). 12
Table 2.3 Effects of 1,25(OH)2D sufficiency on several systems and diseases, summarized from (Battault et al., 2013). 17

Table 3.1 Octanol:water PCs for vitamin D and its metabolites 33
Table 3.2 Parameters for adult humans and corresponding significance in the PBPK model. Parameters adapted from (Williams and Leggett, 1989), with the exception of volume percent of adipose tissue, which is given by Equation (3.3). 36
Table 3.3 Estimation of conversion factors for V_{max}. [1]mg/g tissue, [2]nmol/g protein. [a]Rajwade et al. (1975), [b]Clement et al. (2005), [c]Wilson et al. (2003), [d]Hrycay and Bandiera (2007), [e]Ghazarian et al. (1974). 39

Table 4.1 A comparison of identifiability and estimability (summarized from Table 1 in McLean et al. (2012)) 59
Table 4.2 Sensitivity coefficients for **S**, ordered by relative sensitivity. The horizontal line separates identifiable from unidentifiable parameters. 72

Table 5.1 Final optimized parameter values. Bold-face entries with an asterisk indicate optimized parameters, with *Lower* and *Upper* denoting the 95% confidence intervals (CI) on an optimized parameter. 86

Table A.1 Compound and enzyme abbreviations. 100
Table A.2 Compartmental abbreviations 101
Table A.3 Table of primary symbols 102
Table A.4 Kinetic parameter ranges and literature sources. *Units are pmol/min/mg p450. 103
Table A.5 Initial (total and free) concentration of metabolites in the venous compartment. 104

LIST OF FIGURES

Figure 2.1 Carbon-numbered structures of vitamin D2 and vitamin D3. Note the red circled area on the D2 molecule indicates the difference in the two structures. 6
Figure 2.2 Hydroxylation of vitamin D at the 25 carbon, converting the molecule to 25(OH)D, the first metabolite in the vitamin D cascade. 8
Figure 2.3 Metabolic cascade of vitamin D. Circled areas indicated hydroxylation sites. Enzymes involved in the metabolism are listed in the figure. 9

Figure 3.1 Example of a PBPK model 25
Figure 3.2 Initial PBPK model 27
Figure 3.3 Schematic of the vitamin D metabolism cascade. 28
Figure 3.4 PBPK total model 29
Figure 3.5 Example of $P_{Adi}(t)$ for 25(OH)D 37
Figure 3.6 Sigmoidal curves representing Equations (3.14) and (3.15) 41

Figure 4.1 Comparison of sensitivity of parameters over time. Kinetic parameters for CYP2R1 are shown on the left and α_1, α_2 on the right. 69
Figure 4.2 Total sensitivity ranking. Note the logarithmic scale for the sensitivity coefficients. Numbers above data points indicate the parameter. 71
Figure 4.3 Ranked parameters utilizing the spectrum of the information matrix. Parameters that are both estimable and identifiable are circled in red. 75

Figure 5.1 Initial model results. This includes only the primary and secondary pathways. 81
Figure 5.2 Time-course display of the values of the dynamic adipose partition coefficients. 82
Figure 5.3 Model results with the addition of dynamic adipose partition coefficients. 84
Figure 5.4 Model results with the addition of dynamic adipose partition coefficients as well as dynamic enzyme amounts. 85

Figure B.1 Comparison of sensitivity calculations. 111
Figure B.2 Various transformations of the working simplex. 117
Figure B.3 Compass Search step directions. The blue compass indicates the location of the previous iteration. 119

Figure C.1 Comparison of model fit to the two-pathway (left) and two-site (right) metabolic hypotheses. Figures adapted from (Cuello et al., 2012). 124
Figure C.2 Time-course sensitivities for fixed parameter values. Adapted from (Cuello et al., 2012) 125
Figure C.3 3D sensitivity analysis allowing V_{max} to vary in the two-pathway model. Adapted from (Cuello et al., 2012). 127
Figure C.4 *In vitro* optimization results for lindane. Adapted from (Sawyer et al., in review). 128
Figure C.5 *In vivo* optimization results for lindane. Adapted from (Sawyer et al., in review). 129

Chapter 1: Introduction

One only has to do a brief search in the news to find evidence that vitamin D is a hot topic. Nicknamed the "sunlight vitamin" because of its synthesis from UV exposure, vitamin D has been linked to everything from cardiovascular health to respiratory wellness to prevention of obesity (Vitamin D Council, 2013). As the Industrial Revolution led to an increasing amount of the workday spent indoors, adults and children alike began to show symptoms of vitamin D deficiency. In addition, This trend has continued into present day: With the advent of new technology such as air conditioning and entertainment systems, it is suggested that the average American spends 93% of every 24-hour period indoors (Wagner et al., 2008). This change in lifestyle, and by proxy sunlight exposure, has individuals increasingly turning to over-the-counter vitamin D supplements to augment what little sunlight exposure they get. Despite the free availability of sunlight, almost 33% of the US population is either deficient or at risk for vitamin D deficiency (Looker et al., 2011) and similar results can be seen worldwide (Holick, 2007; Lips, 2010). The question is now raised: How much supplementation is necessary to maintain adequate

levels of vitamin D?

Every individual will have a distinct response to a dose of medication because of differences in absorption, distribution, metabolism, and excretion (ADME) of the substance. However, the behavior of a compound, such as vitamin D, can be described for a general population using physiologically-based pharmacokinetic (PBPK) models. PBPK modeling is a mathematically-based modeling technique that takes physiological constraints–such as tissue volume, blood flows, and metabolism rates–and couples these parameters with chemical properties of the compound of interest. With this powerful combination, PBPK models can predict ADME information for a general–or a specific–population, including response to different exposure routes and dose levels.

With such modeling techniques available at our fingertips, it is expected that a mathematical model with the capability to predict ADME information exists for vitamin D; however, this is not the case. The majority of mathematical models that exist for vitamin D focus on the interplay between vitamin D (and/or its metabolites) and other compounds (Dunn, 1988; Chun et al., 2012), either look at seasonal variation due to sunlight exposure (Holick, 1987; Diffey, 2010; Krzyścin et al., 2011), or are models that look specifically at the kinetics of solar absorption, and tangentially, vitamin D (Olds et al., 2008). No PBPK model of the vitamin D metabolic cascade is available.

This work constructs a PBPK model for vitamin D, with a focus on oral supplementation of vitamin D without sunlight exposure. Several studies exist that restrict sunlight exposure as part of experimental design; we focus on an experiment that investigates a daily dose of vitamin D at common supplementation levels (Fisk et al., 2012). By choosing this type of study design, we can more accurately reflect the behavior of a typical individual with the constructed PBPK model.

This dissertation is organized as follows: In Chapter 2, we introduce the essential

biological background of vitamin D and its metabolic cascade. A significant portion of this chapter is devoted to the enzymes involved in the pathway; understanding the complexities of the pathway is the key to determining appropriate changes for the model. Chapter 3 leads us into a description of physiologically-based pharmacokinetic (PBPK) modeling techniques and these techniques applied to the vitamin D cascade. We first detail the basic PBPK model before introducing several adaptions based on observed biological phenomena, such as the lack of steady-state compartment conditions for the adipose tissue partition coefficients. We then demonstrate the existence and uniqueness of solutions to the full PBPK model. In Chapter 4, we describe the methodologies for model validation and calibration; these discussions include background on identifiability and estimability, and includes our method for parameter estimation. This method includes a discussion of sensitivity analysis and subset selection which will determine the parameters than can be estimated from data. We present the results from the parameter estimation procedure in Chapter 5; within this chapter, we also demonstrate and justify the effects of each model adaptation. Future research directions and conclusions can be found in Chapter 6.

A series of appendices is also included. Appendix A contains relevant tables not present within the main body of this work, as well as a description of how we calculated our initial conditions for the model. The calculation technique is novel, and specific to the vitamin D metabolic cascade. Within Appendix B are discussions of the approximation and optimization techniques used throughout this work. Finally, Appendix C contains a summary of previous work on PBPK models. Each section in Appendix C is taken from papers written in conjunction with undergraduate students in summer research experiences.

Chapter 2: Biology and Kinetics of Vitamin D

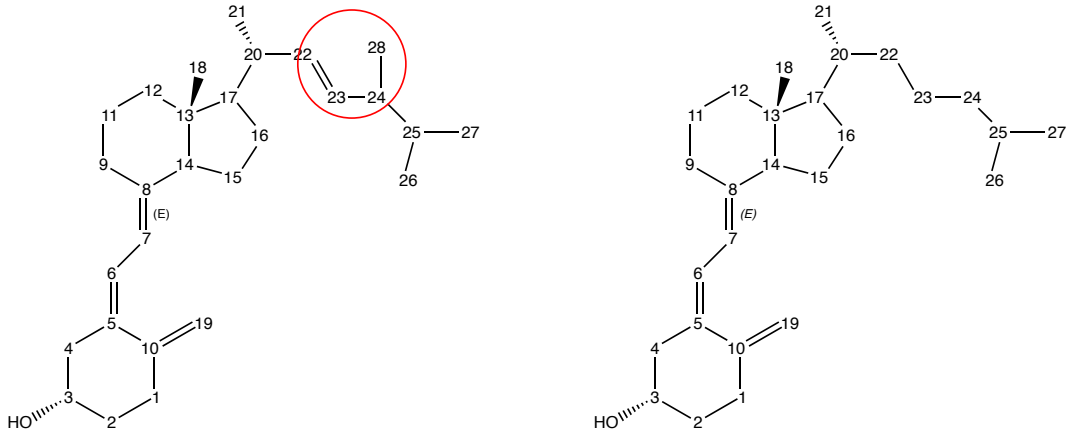
Vitamin D and its metabolites serve a variety of regulatory functions in the body. From regulating the sequestering of calcium to implications of the prevention of autoimmune diseases, vitamin D plays an important role in daily function. This chapter discusses the biology of vitamin D and the metabolites that we are interested in as well as the kinetics behind the enzymes that are involved in the metabolic process. Each metabolite has a particular function in the cascade, from supply to storage to regulation and activation, and will be discussed in detail. Enzymes involved in the cascade range from single-function enzymes, capable of one type of hydroxylation activity, to multi-function enzymes with the capability to produce a variety of metabolites. This chapter will also discuss the biology behind the activation and regulation of the metabolic process, with emphasis on the regulatory functions of the vitamin D metabolite, 1,25(OH)2D.

2.1 The Biology

With the onset of the Industrial Revolution, the average amount of sunlight exposure a typical urban individual received decreased drastically. As a result, incidences of rickets–characterized in children by severe growth retardation and bowing/bending of the bones, and in adults by osteomalacia–skyrocketed to the point where an estimated 90% of children in urban city centers had manifestations of rickets. The late 19th century brought the hypothesis that lack of sunlight was to blame, and experimental evidence verified that exposure to ultraviolet radiation resulted in a cure for rickets. Further experimentation led to identification of vitamin D as the hormone involved in the process. The interested reader is referred to (Holick, 2003; Wolf, 2004) for further perspective on the history of vitamin D.

There are two types of vitamin D: ergocalciferol (D2) and cholecalciferol (D3). The carbon-labeled molecular structures are shown in Figure 2.1; the difference between the two molecules is circled in Figure 2.1a, occurring at the 24 carbon. Ergocalciferol is derived through conversion of a precursor molecule in plants by ultraviolet (UV-B) radiation. Cholecalciferol is produced in animals in the same fashion. There are several hypothesized differences in binding affinities to transport proteins and metabolic enzymes due to the distinct chemical structure. Indeed, some studies have suggested that D2 is not as effective as a supplement as D3 (Houghton and Vieth, 2006), whereas Holick et al. (2008) has suggested no difference in the absorption and utilization of D2 and D3. A review of studies on the differences (or lack thereof) in metabolism of D2 and D3 can be found in Tripkovic et al. (2012). However, at the tissue level, changes in the metabolism of D2 and D3 are considered minor in that the biologic activities of the primary metabolite for both D2 and D3 are comparable (Bikle, 2009). We will follow the convention of not distinguishing

between the two forms of vitamin D unless necessary; to simplify the discussion about the metabolic kinetics and to reflect the data that is used to calibrate the model, we will defer to D3 as the primary form of vitamin D. Thus, nomenclature for vitamin D and its metabolites will follow the naming convention for D3.



(a) Ergocalciferol (D2) (b) Cholecalciferol (D3)

Figure 2.1: Carbon-numbered structures of vitamin D2 and vitamin D3. Note the red circled area on the D2 molecule indicates the difference in the two structures.

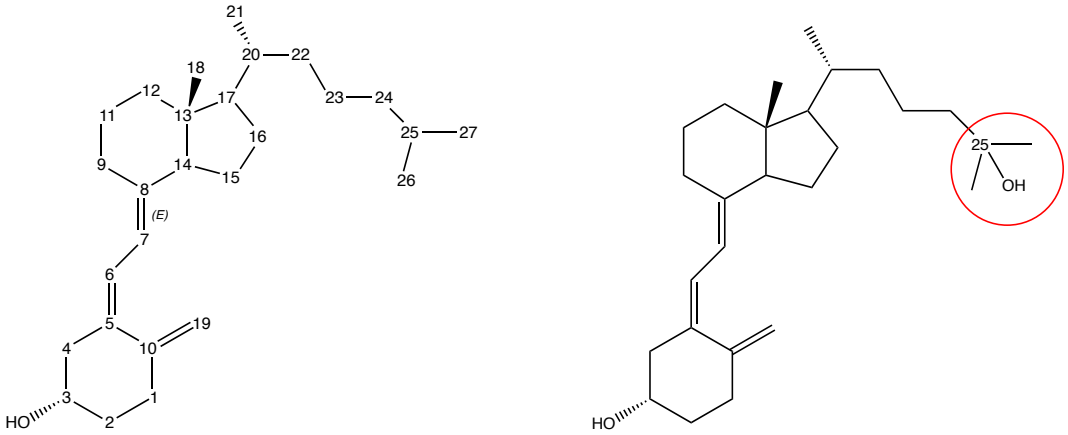
Vitamin D plays a significant role in regulating calcium levels in the body, specifically through actions of the vitamin D receptor (VDR) in target tissues. Levels of vitamin D metabolites trigger VDR-regulated genes in the gut and bones to either release or absorb calcium (Adams and Hewison, 2010). This direct correlation between vitamin D and calcium makes it no surprise that the primary function of vitamin D is musculoskeletal regulation. Deficiencies in vitamin D levels (and subsequently, its metabolites) lead to rachitic symptoms in the form of "soft" bones and muscle weakness. Vitamin D is also implicated in a variety of other systems, including the endocrine, immune, and cardiovascular systems. Recent studies have brought vitamin D to light as a significant

player in the generation of antimicrobial components, cellular function regulation, and in the implication of prevention of diabetes and multiple sclerosis (Dusso et al., 2005; Adams and Hewison, 2010; Zhu and DeLuca, 2012; Morán-Auth et al., 2013).

2.2 Kinetics

The primary action of an enzyme on vitamin D or one of its metabolites is the hydroxylation– the addition of an oxygen-hydrogen (OH) side chain–of various carbon molecules. The location of the new OH chain is given by the name of the metabolite; for example, 25(OH)D is the vitamin D molecule with a OH side chain added at the 25 carbon site, as shown in Figure 2.2. Multiple hydroxylation points are distinguished by a comma in the name of the metabolite, so 1,25(OH)2D has two OH side chains at the 1 and 25 carbons whereas 24,25(OH)2D has the side chains at the 24 and 25 carbons. It is important to note that some references use the naming scheme of "1α" to denote the carbon at the 1 position. The α indicates that the carbon mentioned is the first that attaches to the functional group. Similarly, β denotes the second carbon attached. We will simplify our notation and leave out the attachment order unless necessary.

The biologically-active compound in the vitamin D cascade is not vitamin D itself, but rather one of its metabolites. Cholecalciferol is often considered to be the starting point for the vitamin D cascade. This hormone is converted to calcifediol, also known as 25(OH)D, in the first metabolic step. 25(OH)D is then converted to either the active form calciferol (1,25(OH)2D), or to irrelevant metabolites not included in this model. 1,25(OH)2D is subsequently inactivated through a third step. This pathway, as well as the names of the cytochrome P450 (CYP) enzymes involved in the pathway, is shown in Figure 2.3. Hydroxyl chains that distinguish a metabolite from its parent compound are



(a) Vitamin D (b) 25(OH)D

Figure 2.2: Hydroxylation of vitamin D at the 25 carbon, converting the molecule to 25(OH)D, the first metabolite in the vitamin D cascade.

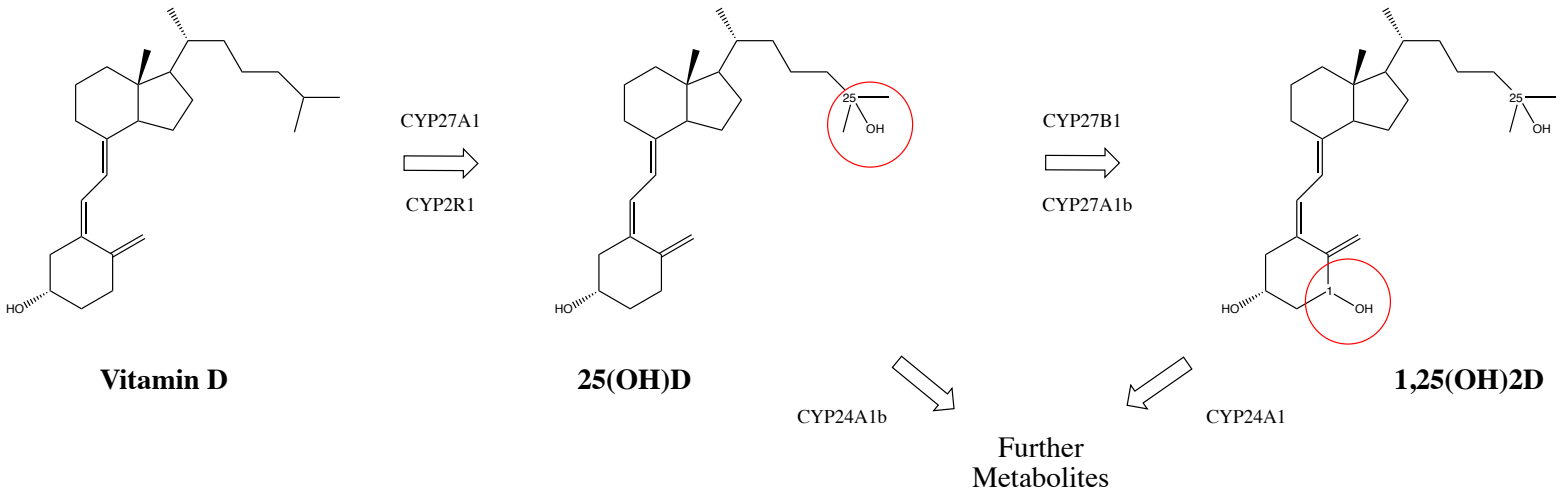


Figure 2.3: Metabolic cascade of vitamin D. Circled areas indicated hydroxylation sites. Enzymes involved in the metabolism are listed in the figure.

circled at each step.

There are a variety of compounds that have vitamin D in their metabolic lineage; we have chosen to include only vitamin D and two of its metabolites. Table 2.1 indicates the reasons for our focus on these compounds.

Table 2.1: Primary reasons for compound choices

Compound	Reason for Inclusion
Vitamin D	Primary source for metabolic cascade; dosing source for model data obtained through diet and sunlight exposure.
25(OH)D	Storage form of vitamin D; typical model data is given in terms of concentration of 25(OH)D.
1,25(OH)2D	Biologically-active form of vitamin D; involved in regulation of metabolic pathway

2.2.1 Initial Activation

There are two typical routes of entry for vitamin D. The first (and primary) route is through synthesis of the hormone in the skin. Vitamin D is created through the action of UV-B radiation which controls the conversion of a precursor molecule, 7-dehydrocholesterol (DHC), in the skin to pre-vitamin D, which is subsequently converted to cholecalciferol through heat. This step is tightly regulated by the local concentrations of vitamin D; if the local concentration of cholecalciferol is high enough, DHC will be converted to non-active products to prevent overdosing of vitamin D. This on-site regulation of the skin production of vitamin D has been shown by Holick et al. (2008) and others and suggests it is virtually impossible to overdose on vitamin D strictly through skin production of the

hormone. Early studies on the effect of sunlight on vitamin D have shown that synthesis of D3 from DHC plateaus at approximately 15% of the original amount of DHC (Holick, 1987). If cholecalciferol is produced, it becomes tightly bound to the vitamin D binding protein (DBP) and enters the bloodstream where it is slowly released into the body and metabolized.

A second route of entry for vitamin D is through diet. When vitamin D is obtained through oral supplementation, it is transported along with chylomicrons and lipoproteins before being released to DBP. Some of the vitamin D remains on the chylomicron remnants and is excreted; Vieth (2005) estimates that up to 75% of orally-obtained vitamin D is excreted without being absorbed into tissues or metabolized. Fraser (1983) has proposed that vitamin D introduced to the liver on chylomicron fragments may be preferentially directed toward metabolic pathways that inactive the compound. This loss of compound or function has significant implication on utilization of the oral dose.

In contrast, skin-synthesized vitamin D is presented to the body primarily bound to various proteins such as albumin and DBP. The strength of the affinity to these proteins has a significant effect on the biological activity of vitamin D and each of its metabolites. The majority ($> 99\%$) of vitamin D and its metabolites are bound to proteins in the plasma, leaving very little "free". Consequentially, the half-lives of vitamin D and its metabolites are significantly longer compared to other steroid hormones. This, in conjunction with the implication that the lipid solubility of the compound allows for rapid transport across membranes, suggests that uptake of all of these metabolites by organs is elimination-limited. If a compound has a very small free percentage, a long plasma half-life, and uptake by tissues is elimination-limited, there is a very strong probability that the hormone behaves according to the *Free Hormone Hypothesis* (FFH) (Mendel, 1992).

Definition 2.2.1 (Free Hormone Hypothesis)**.** *The biological activity of a hormone is affected by the unbound (free) rather than the protein-bound concentration of hormone in the plasma, as measured at equilibrium in vitro. Equivalently, the concentration of hormone within tissues is affected by the free rather than the protein-bound concentration of hormone in the plasma. (Recant and Riggs, 1952)*

The FFH is expected to be valid if uptake of a hormone by a tissue is slow compared to the rates of intracapillary events (plasma flow and protein-hormone interactions, for example). Hormones that are eliminated slowly are also more likely candidates for application of the FFH. Metabolites of vitamin D, especially 25(OH)D, fit the requirements for the FFH: The half-life of 25(OH)D in the plasma is on the order of two weeks and the high affinity between the 25(OH)D molecule and DBP severely limits dissociation of the complex (Bouillon, 2011).

Vitamin D and its metabolites primarily bind to DBP with tight affinities. To estimate the free percent of vitamin D and its metabolites, we can use techniques discussed in Feldman (1972). Table 2.2 shows the affinity to DBP and the free percent for vitamin D, 25(OH)D, and 1,25(OH)2D; the equations used to calculate the free percentages are discussed in Appendix A.2.

Table 2.2: Percent free of each substrate and the affinity of each to DBP. [1]As calculated using methods described in Feldman (1972).

	Vitamin D	25(OH)D	1,25(OH)2D	Source
% Free[1]	0.07	0.04	0.4	[1]Calculated
Half-life	24 hrs	3 weeks	4 hrs	Zerwekh (2008)
DBP Affinity	$4 \times 10^8\ M^{-1}$	$7 \times 10^8\ M^{-1}$	$4 \times 10^7\ M^{-1}$	Chun et al. (2012)

The FFH may not hold if significant changes to the concentration of either the hormone or its binding compound occur. However, circulating levels of DBP are much higher than the circulating levels of the metabolites, and the normal occupancy of the metabolite binding site of DBP is less than 5%; this indicates a large surplus of binding sites available if concentrations of a metabolite increase (Bouillon, 2011). We incorporate the FFH into our model through regulation of the biological availability of the metabolites. Further details about use of the hypothesis will be discussed in Appendix A.2.

2.2.2 25-Hydroxylation

The first step in the activation of vitamin D is the conversion to 25(OH)D. This metabolite is important in the cascade because it serves as the primary storage form in the vitamin D pathway. Levels of circulating vitamin D are highly dependent upon recent sun exposure and ingestion of vitamin D because of the fairly rapid half-life (24 hours); in contrast, the tight affinity of 25(OH)D with DBP and other plasma transport proteins, levels of circulating 25(OH)D are slow to change and reflect vitamin D stores over a longer time period. This makes 25(OH)D the ideal metabolite to measure for levels of sufficiency (Zerwekh, 2008).

Conversion of vitamin D to 25(OH)D occurs mainly in the liver through the activities of two cytochrome P450 (CYP) enzymes, CYP27A1 and CYP2R1. There are several other CYPs that have the potential to 25-hydroxylate vitamin D in humans, including microsomal CYPs 3A4 and 2J2/3 (Prosser and Jones, 2004); however, evidence suggests that CYP3A4 acts preferentially on a variation of vitamin D, namely the metabolite 1α(OH)D, and does not show activity toward vitamin D itself (Gupta et al., 2004). Although CYP2J3 shows strong activity toward vitamin D and 1α(OH)D in rats, the

human analog, CYP2J2, shows activity 10-fold times weaker than its counterpart and the affinity between CYP2J2 and substrates appears much too low to be of physiological consequence (Aiba et al., 2006). We thus chose to focus on CYPs 27A1 and 2R1 because of their demonstrated kinetic behavior at physiological levels of vitamin D.

CYP27A1

The enzyme CYP27A1 was the first to be associated primarily with the creation of 25(OH)D in rabbit and rat liver mitochondria. Identification of human kinetic parameters for 25-hydroxylation of vitamin D using expression in *E. coli* was completed by Sawada et al. (2000), who also demonstrated that CYP27A1 has the capacity to hydroxylate vitamin D in several other locations. Although cells transfected with CYP27A1 have demonstrated capacity to synthesize 25(OH)D, several factors suggest that CYP27A1 is not the sole nor even the major player in the 25-hydroxylation of vitamin D:

I. Although CYP27A1 is able to hydroxylate vitamin D, the enzyme shows preferential activity toward other hydroxylated forms, such as 1α(OH)D and 25(OH)D (Sawada et al., 2000; Shinkyo et al., 2004). To distinguish between activity of CYP27A1 on vitamin D versus on 25(OH)D, we will use the notation "CYP27A1b" to indicate activity of CYP27A1 on 25(OH)D.

II. CYP27A1 is a low-affinity, high capacity mitochondrial enzyme. CYP27A1 also has K_m values (the concentration at which one-half of the maximum velocity of enzyme activity is achieved) at least 30-fold higher than typical circulating concentrations of the vitamin D substrate, implying that at physiological concentrations, CYP27A1 is not the primary enzyme involved in synthesis of 25(OH)D (Zhu et al., 2010).

III. In rat livers, Fukushima et al. (1975) demonstrated that there are two separate

types of 25-hydroxylase enzyme activities: a high-affinity, low-capacity form, and a low-affinity, high-capacity form (presumably CYP27A1).

IV. Finally, 25-hydroxylation of vitamin D2 by CYP27A1 has not been detected (Guo et al., 1993), although mammals have been shown to be able to utilize D2 as a supplemental source (Jones et al., 1998).

Together, these statements suggest that CYP27A1 is not the primary enzyme involved in the metabolism of native vitamin D. However, because of its low-affinity, CYP27A1 may become more relevant at pharmacological concentrations of vitamin D. We therefore consider the possibility that CYP27A1 is relegated to a multi-purpose enzyme with minor contributions toward 25(OH)D synthesis.

CYP2R1

With CYP27A1 demoted as a minor player in the synthesis of 25(OH)D, other CYPs were investigated for synthesis potential. Early kinetic studies had identified a potential microsomal protein involved in the production of 25(OH)D; however, it took until 2003 for identification of this enzyme as CYP2R1, which was previously an "orphan" enzyme with no known biological function (Cheng et al., 2003, 2004). In contrast to CYP27A1, CYP2R1 fits the activity profile predicted above: CYP2R1 has the ability to 25-hydroxylate D2 and D3 with equal efficiency as well as a lack of preference toward 1α-hydroxylated vitamin D compounds; CYP2R1 is a microsomal enzyme with a high-affinity, low-capacity form; and CYP2R1 does not act on 25(OH)D (Sawada et al., 2000; Shinkyo et al., 2004; Jones, 2010; Zhu and DeLuca, 2012).

2.2.3 1-Hydroxylation

The second step in the activation of vitamin D is through the hydroxylation of 25(OH)D in the 1 position to form the biologically-active form of vitamin D, 1,25(OH)2D. This metabolite has a short half-life compared to its precursor, 25(OH)D. This is due to the utilization of 1,25(OH)2D by several different receptors, but primarily through the actions of the vitamin D receptor, VDR. The affinity of DBP toward 1,25(OH)2D is roughly an order of magnitude less than the affinity of DBP toward 25(OH)D; this frees up more of the metabolite for biological activity mediated by VDR.

The main location for conversion of 25(OH)D to 1,25(OH)2D is in the proximal tubule in the kidney and utilizes the enzyme CYP27B1. After conversion, 1,25(OH)2D enters circulation where, in conjunction with VDR, it acts on organs and cells in a hormone-like manner. There are two primary functions of circulating 1,25(OH)2D in the human body. The first is to increase the efficiency of intestinal calcium and phosphorous absorption, and the second is to control bone growth. Secondary roles include regulation of renin (an enzyme controlling arterial blood pressure) production in the kidney and stimulation of insulin secretion in the pancreas (Holick, 2007).

Although the primary site of conversion of 25(OH)D into 1,25(OH)2D is in the kidney, there are a number of extra-renal sites, such as the muscles, prostate, immune cells, and pancreas, that express CYP27B1 and can supply local need for 1,25(OH)2D. Effects of 1,25(OH)2D on classical and nonclassical tissues are described in Table 2.3.

CYP27B1

CYP27B1 is primarily found in the kidneys, although several studies have indicated extra-renal sites for this enzyme, including immune cells and keratinocytes (Jones, 2008;

Table 2.3: Effects of 1,25(OH)2D sufficiency on several systems and diseases, summarized from (Battault et al., 2013).

System/disease	Effect
Skeletal	Regulation of intestinal calcium and phosphorous absorption. Promotion of development of osteoclasts, which release calcium and phosphate into general circulation.
Muscle	Enhanced calcium homeostasis regulation and balancing of hyperparathyroidism.
Cardiovascular	(Indirect) modulation of blood pressure.
Immune	Enhancement of antimicrobial properties of immune cells. Mitigation of adaptive immunity
Cancer	Suppression of proliferation and stimulation of differentiation of cancer cells. Regulation of apoptosis (cell death) in cancer cells.
Diabetes	Influence on pancreatic beta cell function and improvement of insulin sensitivity in target cells.

Van Etten et al., 2008; Henry, 2011). CYP27B1 acts by hydroxylating 25(OH)D at the 1 carbon, producing the active form of vitamin D. This process is tightly regulated and is both induced and inhibited by various factors. 1,25(OH)2D itself is a negative regulator of CYP27B1 transcription (and therefore activity), whereas levels of parathyroid gland hormone (PTH) and calcium stimulate CYP27B1 activity (Van Etten et al., 2008). Different extra-renal tissues show varying regulation of CYP27B1 activity: Vitamin D and UV-B levels upregulate CYP27B1 activity in keratinocytes, but 1,25(OH)2D (important for regulation in renal cells) has no effect on CYP27B1 activity in bone cells. Several new studies are demonstrating the role of CYP27B1 in the immune system through its local production of 1,25(OH)2D (Van Etten et al., 2008; White, 2012; Morán-Auth et al., 2013).

Minor Pathways

Formation of 1,25(OH)2D can also occur via CYP27A1. As stated earlier, CYP27A1 has the capability to metabolize a variety of substrates; one of these reactions is the 1-hydroxylation of 25(OH)D. However, the reaction speed of the 1 pathway is ten times slower compared to 25-hydroxylation of vitamin D. For the purposes of this model, we will still consider the actions of CYP27A1 on 25(OH)D valid since other reaction values for 1-hydroxylation are similar to that of the same reaction catalyzed by CYP27B1 (Sakaki et al., 2005).

2.2.4 Deactivation of Vitamin D

Due to its importance in many different biological functions, 1,25(OH)2D must be tightly regulated. Too much calciferol in the body can lead to hypercalcemia, where calcium levels in the blood are elevated due to excessive skeletal release of calcium. Too little calciferol has the opposite effect, where calcium levels in the blood are low due to absorption of calcium into the skeletal system. Fortunately, 1,25(OH)2D can both stimulate and inhibit its production and destruction. As discussed earlier, CYP27B1 is activated in the presence of low amounts of calciferol; similar enzymes that are involved in the deactivation of 1,25(OH)2D are inhibited at low concentrations of calciferol, but stimulated at high concentrations. The most common regulatory enzyme of 1,25(OH)2D is CYP24A1, although several other enzymes have the capability to utilize 1,25(OH)2D as a substrate but at much lower reaction rates (e.g. CYP3A4 (Xu et al., 2006)).

CYP24A1

CYP24A1 is, as are several other hydroxylases in the vitamin D cascade, a multi-catalytic enzyme. Specifically, CYP24A1 can act on 25(OH)D and 1,25(OH)2D by 24-hydroxylating the substrate and ultimately deactivating the compound. As is the case for CYP27A1, we will denote the actions of CYP24A1 on 25(OH)D as "CYP24A1b" for convenience in later discussions. CYP24A1 is found throughout the body in a wide variety of target tissues that contain the VDR molecule (Jones et al., 2012). The kidney, however, is the producer of circulating levels of 24,25(OH)2D; thus for this model, we consider the activities of CYP24A1 only in the renal system.

CYP24A1 is strongly inducible by VDR and 1,25(OH)2D; high levels of 1,25(OH)2D activate the enzyme and lead to catabolism of calciferol and protect cells from excess VDR pathway activation. In addition, CYP24A1 balances the production of 1,25(OH)2D through acting in conjunction with CYP27B1 and utilizing 25(OH)D as a substrate.

2.2.5 Regulation of the Vitamin D Pathway

The initial regulatory step for the activation of vitamin D occurs in the synthesis of the pre-vitamin D from DHC. When skin is exposed to sunlight, a tanning effect occurs through an increase in melanin content. This down-regulates production of vitamin D since melanin competes with DHC for absorption of UV-B radiation (Battault et al., 2013). Once vitamin D is synthesized (or ingested), conversion to 25(OH)D occurs without much moderation. However, metabolism of 25(OH)D to 1,25(OH)2D is tightly regulated, with a series of alternate pathways available for deactivation of vitamin D.

A novel pathway for regulation of vitamin D is the 4β-hydroxylation of 25(OH)D by CYP3A4, recently discovered by Wang et al. (2012). In addition, CYP3A4 has the

capability to hydroxylate 1,25(OH)2D in several positions; however, metabolic clearance of CYP3A4 was less than that of CYP24A1 except in the intestine and liver (Xu et al., 2006). Since the behavior of CYP3A4 is not as well understood as that of CYP24A1 in the kidneys, we chose to omit CYP3A4 from our model.

Regulation of the entire vitamin D pathway revolves around the regulation of circulating and local concentrations of 1,25(OH)2D. It is understood that 1,25(OH)2D asserts its effects through VDR; therefore, by regulating the amount of ligand available to the VDR, cellular effects can be modulated. The easiest way of regulating concentrations of a substance is to control its synthesis and destruction.

The main enzyme responsible for the synthesis of 1,25(OH)2D is CYP27B1. This enzyme is stimulated by a variety of factors, including low circulating levels of calcium and phosphorous, the parathyroid hormone (PTH), and by fibroblast growth factor 23 (FGF23). In contrast, CYP27B1 is down-regulated by its own creation - 1,25(OH)2D. The compound 1,25(OH)2D down-regulates CYP27B1 not through inhibition of the substrate pocket in the enzyme, but by direct control over the transcription through interaction with VDR (Henry, 2011). By suppressing transcription, the physical amount of CYP27B1 is decreased.

The rate of catabolism of 1,25(OH)2D is controlled through several factors, but primarily through control over levels of CYP24A1. CYP24A1 is induced by FGF23 in various cell types, but is suppressed by PTH in the kidneys (Jones et al., 2012). Levels of 1,25(OH)2D increase the transcription of CYP24A1. CYP24A1 can either act on 1,25(OH)2D directly, or on 25(OH)D, limiting the initial production of 1,25(OH)2D.

Although the regulatory pathways for the synthesis and catabolism of 1,25(OH)2D include a variety of compounds including FGF23 and PTH, we will focus our model only on the effects of 1,25(OH)2D on the transcription of CYPs 27B1 and 24A1. The effects

of 1,25(OH)2D are often inversely proportional with those of FGF23 and PTH (Henry, 2011), so we are confident that this simplification of the model does not significantly affect the predictive capabilities of the pathway. Detail about how transcriptional regulation is introduced into the model can be found in Chapter 3.

2.3 Summary

Traversing the pathway from activation of the 7-dehydrocholesterol compound to regulation of 1,25(OH)2D is a tightly controlled process. The main enzymes involved–CYPs 27A1, 27B1, 2R1, and 24A1–each have a delicate role in the cascade. The metabolites themselves also must behave according to various binding and regulating factors, and several feedback loops influence the pathway as well. The resulting dance between activation and destruction of vitamin D has implications in a wide range of systems and diseases, from maintaining skeletal health to prevention of autoimmune diseases through control of cellular activities.

Chapter 3:

Model Development and Basic Analysis

Qualitative modeling of vitamin D is not new; almost as soon as vitamin D and its antiricketic properties were discovered, scientists began testing various hypotheses about levels of vitamin D in fish (Bills, 1927), synthesis of the hormone in skin (Hume et al., 1927), and the effects of vitamin D on calcium and phosphate levels in the blood (Havard and Hoyle, 1928). The early 1960's brought about studies that were more concerned with quantitative information about vitamin D. As an example, Chandler and Cragle (1962) were two of the first researchers to quantify the relationship between calcium, phosphorus and vitamin D using mathematical techniques. As available technology improved, so did the complexity of the models. Improvements in experimental methods and discovery of specific metabolites raised different questions about the quantitative behavior of vitamin D and its newly discovered metabolic children. Scientists became interested in how 25(OH)D could help different populations outside the traditional rachitic base of patients, such as epileptics (Stamp et al., 1972) and those with renal failure (Coburn et al., 1973). More recent models have tended to focus on a specific aspect of vitamin D and the human

body, be it synthesis of vitamin D from sunlight (Diffey, 2010) or the effects of vitamin D sufficiency on calcium regulation and bone formation (Rattanamongkonkul et al., 2011) and immune system support (Hewison, 2010). However, to date, there are no mathematical models that describe the vitamin D cascade from synthesis/absorption to inactivation, much less any models that take a wide variety of physiological parameters into account.

In addition to omitting a majority of descriptive parameters, mathematical functions used to describe behavior of data sets often have very little biological mechanistic justification. For example, Heaney et al. (2003) derive the increasing form of an exponential decay function (Equation (3.1)) to relate the serum concentration of 25(OH)D to the amount of oral vitamin D dosing into the system:

$$C(t) = C(0) + a(1 - e^{-kt}), \tag{3.1}$$

where $C(t)$ is the concentration at time t, a is the increment at equilibrium assuming a constant input of vitamin D, and k is the rate constant representing the total mass of 25(OH)D used per day. The authors admit that this model is a first approximation and describe several simplifications that have reduced the biological accuracy of the model. However, for the purposes of their analysis, Equation (3.1) is sufficient to describe the overall effect of vitamin D supplementation on levels of 25(OH)D in replete (those with sufficient amounts) human subjects.

We are interested in a model that incorporates a myriad of physiological parameters, including body composition and age, and biologically-justifiable equations governing the inter-compartmental transitions and kinetics of vitamin D and its associated metabolites. Fortunately, there are modeling techniques available to combine these pieces, namely in the form of physiologically-based pharmacokinetic models.

3.1 Physiologically-Based Pharmacokinetic Models

A pharmacokinetic model is a quantitative, rather than qualitative description of chemical behavior in biological systems over a period of time. The focus of these systems is typically individual species and the associated chemical breakdown in their tissues. Physiologically-based pharmacokinetic (PBPK) models take ADME principles and combine these with parameters obtained from physiological descriptions of the subjects. PBPK models, in contrast to general pharmacokinetic models, refer to mathematical descriptions of uptake and flow of chemical through tissue groups based on quantitative interactions between biological processes (Krishnan et al., 2010). A brief overview of PBPK models is presented here; the curious reader is referred to Krishnan et al. (2010) for more detail and a software example of PBPK modeling.

PBPK models are based on dividing the organism into compartments. The modeler must choose how to segregate organs or tissue groups based on physiological necessity or experimental interest. There are two approaches on how to choose compartments. The first approach is lumping, where the initial model structure is as detailed as possible, and tissue compartments are combined based on similarity of physiological properties. An example of where this occurs is in the creation of the slowly perfused compartment, where muscle and bone are grouped because of the lower ratio of blood flow to tissue volume, as opposed to, say, that of the brain. Lumped tissues are assumed to be homogeneous with respect to chemical behavior (Krishnan et al., 2010). The second approach is splitting, where the initial model structure is as simple as possible and then compartments are segregated only as necessary to replicate data (Clewell and Clewell III, 2008). Generally speaking, a balance must be struck between lumping and splitting behaviors to generate a model with the minimum degree of complexity necessary to sufficiently explain available

data.

Additions of complexity can come in the form of adding compartments or new mechanisms with associated mathematical representations to generate observed system behavior. Common criteria for selection of compartments include their relevance in metabolism as well as differences in the behavior of solubility of a chemical in a proposed compartment.

3.1.1 Mathematical Description

Compartmental models can be conceptualized as a series of chambers connected through flows (e.g., blood or metabolism). Figure 3.1 illustrates a five compartment model. The

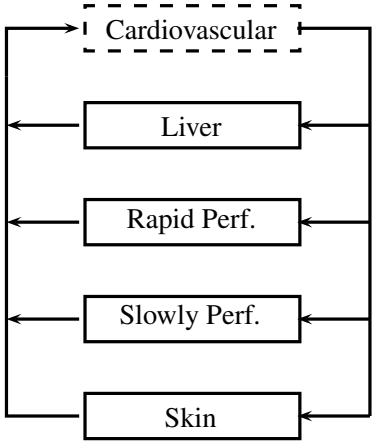


Figure 3.1: Example of a PBPK model

change in the concentration for each compartment can be derived utilizing Fick's first law and the concept of conservation of mass (Krishnan et al., 2010); however, it is sufficient for us to think of the change in concentration as

$$\frac{dA}{dt} = Q(C_{in} - C_{out}) \tag{3.2}$$

where A is the amount in the compartment, Q is the flow through the compartment, and C is the concentration of the chemical coming *in* or *out* of the compartment. Each compartment can be described by an equation of this form. Krishnan et al. (2010) details several modifications that can be made to Equation (3.2) to account for ADME principles.

The resulting mathematical description of a PBPK model is a series of (potentially) nonlinear, coupled ordinary differential equations. There are several computer programs available to solve systems of this form. For the purposes of this project, we utilized the software distribution MATLAB© (version 7.12.0.623 R2011a) and the stiff ODE solver `ode15s`, a differential equation solver available in the standard distribution. Derivations of how differential equation solvers work, particularly related to PBPK models, can be found in (Krishnan et al., 2010).

3.2 Vitamin D Model

3.2.1 General Model

A ten-compartment model (adipose tissue, gastrointestinal (GI) tract, kidney, liver, rapidly perfused tissue, slowly perfused tissue, skin, spleen, and venous and arterial blood) was developed as a base model for each substrate. The skin and GI tract were isolated as compartments of interest due to their functionality as exposure routes. Initially, the spleen compartment was lumped with the GI tract; however, we chose to isolate the spleen due to its unique contribution to blood circulation in the liver. An inherent assumption typically made in PBPK modeling is that each compartment is homogenous; this implies that the concentration of the substance does not change throughout the compartment. A schematic of the PBPK model for the substrate vitamin D is shown in Figure 3.2. We include a

GI lumen compartment in the vitamin D base model that serves as a reservoir for oral doses of vitamin D and allows for passive, nonsaturable, and continuous absorption of vitamin D into the GI tract (Hollander et al., 1978). PBPK models for 25(OH)D and 1,25(OH)2D are similar, differing only in the exclusion of the GI lumen, as we assume that only vitamin D is present in the exposure dose by experimental design. Note that neither vitamin D nor any of its metabolites are volatile; thus, the cardiovascular compartment only serves as a "pass-through" compartment and physiological values for the heart and lungs are incorporated in the rapidly perfused compartment to ensure mass conservation.

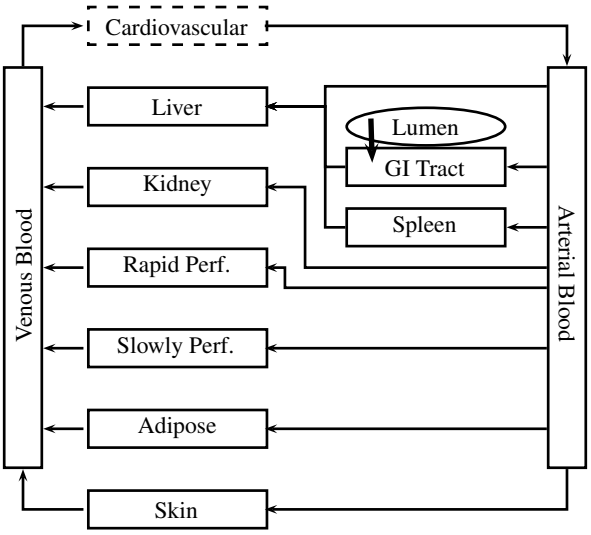


Figure 3.2: Initial PBPK model

Metabolism serves as a continuous flow between the substrate models. The metabolism cascade, with the associated enzymes, is represented in Figure 3.3, and the full PBPK diagram is shown in Figure 3.4.

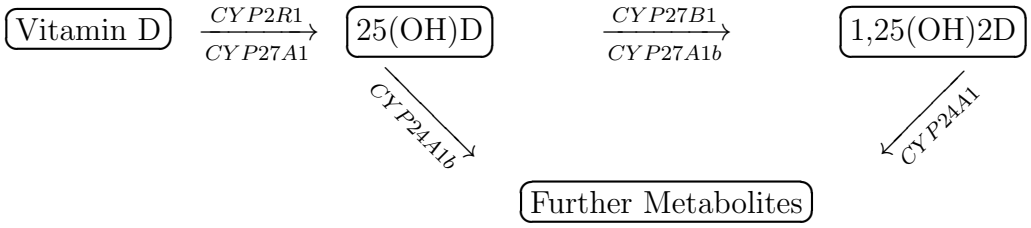


Figure 3.3: Schematic of the vitamin D metabolism cascade.

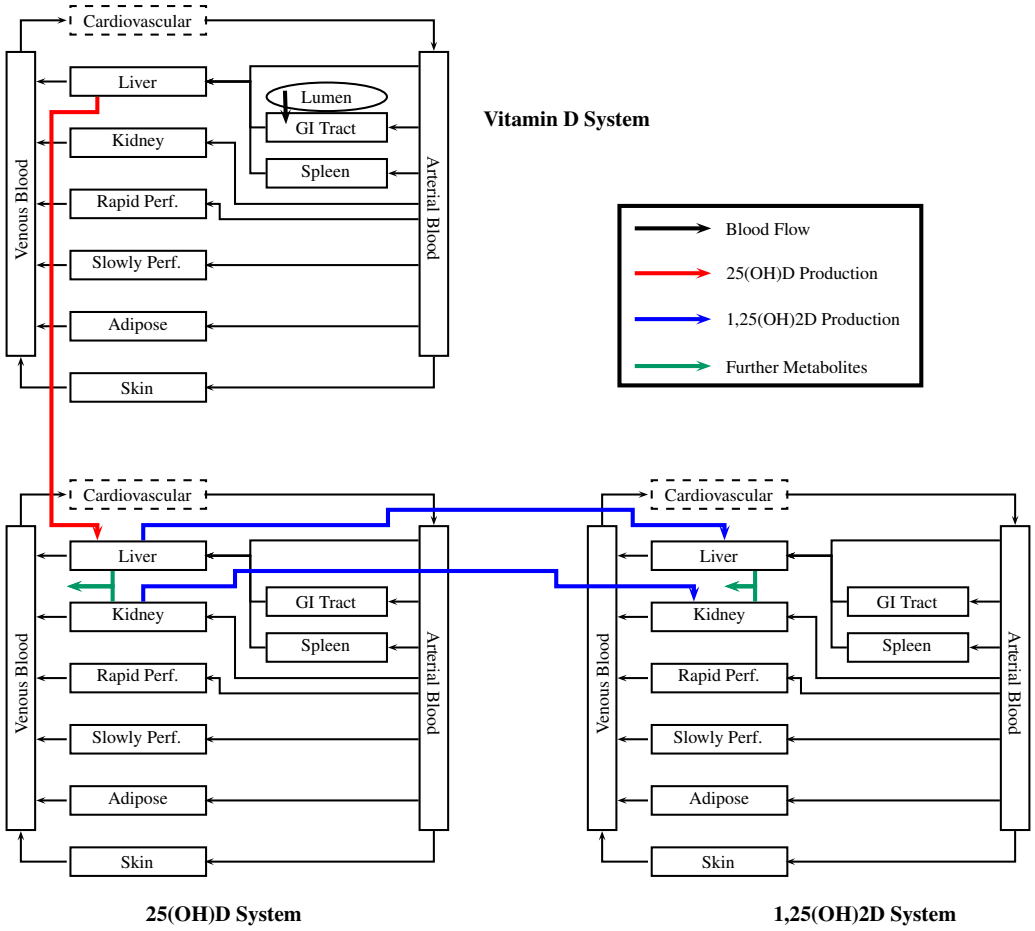


Figure 3.4: PBPK total model

3.2.2 Study Design

Study Selection

As vitamin D can be either orally obtained or internally created, we wanted a dataset that utilized a single exposure route to minimize potential interplay between the routes. Many foods, including milk and cereals, are fortified with vitamin D; this prevalence makes a human experiment with a vitamin D-restricted diet much more difficult to find than one with sunlight restriction. We chose to consider studies that looked at oral dosing under sunlight-restricted conditions, such as studies occurring in winter months or at high latitudes.

Several other considerations went into our choice of datasets. We would like to have the capability to predict values for an average human, male or female, at typical vitamin D supplementation levels. Study data from Fisk et al. (2012) comprises of 40 individuals, approximately equal in division among sex. In addition, the mean statistics provided by the study suggest that all individuals are roughly "average" in age, typical diet intake of vitamin D, and 25(OH)D serum level. The average body-mass index (BMI) for each subject group is slightly higher; we take this into account in parameter selection (Section 3.2.3).

Model Design

Serum 25(OH)D data for parameter estimation was taken from Fisk et al. (2012). The available data was digitized using the MATLAB code `digitization` obtained from the user community website MATLAB Central (Krogstad, 2006). A PBPK model was developed to describe the concentration of vitamin D and its primary metabolites, 25(OH)D and 1,25(OH)2D, in the venous compartment. Vitamin D and its metabolites are present in each individual at the start of the study; thus, it was necessary to estimate the levels of

each substrate in all compartments prior to implementation of the PBPK model. This provides a more accurate initial condition for the model than assuming zero substrate is present, as would be the case in a novel dosing scheme.

Participation criteria indicate that no individual in the study was taking a vitamin D supplement prior to the start of the experiment; this allows for the assumption that levels of each substrates are roughly at a steady-state level. For each dosing scheme, the concentrations of the substrates were estimated using the initial data point for the dose. Explicit details about how compartment steady-state levels were determined can be found in Appendix A.2.

3.2.3 Physiological Parameter Selection

Physiological parameters for volume and blood flow rates for each compartment were obtained from literature. Based on study design, we used the mean of male and female physiological parameters for the volumes and blood flow rates, as given in Williams and Leggett (1989). The rapidly perfused compartment is assumed to be composed of the heart, brain, pancreas, and thyroid. The slowly perfused compartment was assumed to have a blood flow rate fraction equal to $1 - \sum Q_i$ where Q_i is the fractional blood flow rate of all compartments except the arterial and venous compartments. Similarly, the volume fraction of the slowly perfused compartment is assumed to be equal to $1 - \sum V_i$ where V_i is the fractional volume of the remaining compartments. The use of these equations ensures preservation of flow and volume for the total model. The percent of adipose tissue was adjusted to account for the average BMI of the subjects; to convert from BMI to body fat percentage, we used Equation (3.3), taken from (Deurenberg et al., 1991):

$$1.20 \times BMI + 0.23 \times age - 10.8 \times sex - 5.4, \tag{3.3}$$

where $sex = 1$ for males and 0 for females. To account for both sexes present in the study, we used $sex = 0.5$ in Equation (3.3). A table of physiological values can be found in Table 3.2, where the value given is the average of the male and female values taken from (Williams and Leggett, 1989) where body weight is assumed to be 66.5 kg (assuming unit density of tissue) and cardiac output is 375 L/hr.

3.2.4 Partition Coefficients

As chemicals flow though compartments via blood, a percentage of the chemical will become bound to the tissue. This changes the apparent tissue concentration of the chemical. To account for this binding, we use a ratio that relates the binding of one medium to the binding in another, otherwise known as a partition coefficient (PC). A partition coefficient can relate the binding of a chemical in air versus the blood, as might be the case for a volatile chemical, or between tissues and blood. The magnitude of the octanol:water PC of a compound can help determine the lipophilicity of a compound such as non-volatile vitamin D. If the partition coefficient is less than 1, then the chemical is hydrophilic ("water-loving"). As the PC increases, the chemical becomes more lipophilic ("fat-loving"). For an extensive review on the theoretical nature of partition coefficients, the interested reader is referred to (Leo et al., 1971).

Very little is known about the tissue:blood partition coefficients for vitamin D or its metabolites. However, the octanol:water PC is known for each compound, and by utilizing the lipid and water composition of tissues alongside the methods described in Poulin and Krishnan (1995) with adaptions for binding, as given in (Poulin and Theil, 2002), we can estimate the tissue:blood PCs for each compartment. Due to their chemical structure, vitamin D and its metabolites have very high octanol:water PCs, shown in Table 3.1.

To account for this, we will use an adaptation for the adipose:blood PC as described in (Poulin and Haddad, 2012).

Table 3.1: Octanol:water PCs for vitamin D and its metabolites

	Vitamin D	25(OH)D	1,25(OH)2D
$\log P_{o:w}$	10.24	9.14	7.6

The main equations for the calculation of non-adipose tissue:blood PCs, Equation (3.4), and for the adipose:blood PC, Equation (3.5), are found in (Poulin and Theil, 2002):

$$P_{t:b} = \frac{P_{o:w} \times (V_{nlt} + 0.3 \times V_{pht}) + (V_{wt} + 0.7 \times V_{pht})}{P_{o:w} \times (V_{nlb} + 0.3 \times V_{phb}) + (V_{wb} + 0.7 \times V_{phb})} \times K_t \tag{3.4}$$

$$P_{adipose:b} = \frac{P_{vo:w} \times (V_{nlt} + 0.3 \times V_{pht}) + (V_{wt} + 0.7 \times V_{pht})}{P_{vo:w} \times (V_{nlb} + 0.3 \times V_{phb}) + (V_{wb} + 0.7 \times V_{phb})} \times K_{adi}, \tag{3.5}$$

where $P_{o:w}$ is the octanol:water PC, $P_{vo:w}$ is the olive oil:water PC (which can be expressed in terms of the $P_{o:w}$, as shown in Equation (3.6)), V is the fractional tissue volume content for neutral lipids (nl), phospholipids (ph) and water (w), t is tissue, b is blood, and $K_i, i \in \{t, adi\}$ relates the unbound fraction in the blood to the unbound fraction in the non-adipose (t) or adipose (adi) tissue (described in Equation (3.7)).

For highly lipophilic compounds, such as vitamin D and its metabolites, using $P_{o:w}$ to calculate the adipose:blood partition coefficient may result in inaccurate predictions of the volume of distribution at steady state. The reason may be with highly lipophilic compounds, it is the neutral lipid portion, rather than the entire adipose tissue, that accounts for chemical storage. To counter this, Poulin and Haddad (2012) suggest using the

vegetable oil:water (or olive oil:water) PC rather than the octanol:water PC as a parameter in calculating partition coefficients for adipose tissue. For compounds containing at least one oxygen atom in their molecular structure (which is clearly achieved by compounds in the vitamin D cascade), Equation (3.6) gives $P_{vo:w}$ in terms of the more commonly available $P_{o:w}$:

$$\log P_{vo:w} = 1.099 \times \log P_{o:w} - 1.31. \tag{3.6}$$

When calculating partition coefficients, the ratio between the amount of chemical bound in the tissues versus the amount bound in the blood must be accounted for. In Equations (3.4) and (3.5), this ratio is accounted for by the K_t term for non-adipose tissues and K_{adi} for the adipose tissue (De Buck et al., 2007). K_t and K_{adi} are defined in Equation (3.7):

$$K_t = fu_b + .15 \times (1 - fu_b); \qquad K_{adi} = fu_b + .5 \times (1 - fu_b). \tag{3.7}$$

Values for the lipid and water content of each tissue were taken from (Poulin and Theil, 2002). To compensate for the mixed tissue composition of the slowly and rapidly perfused compartments, we used the following adaptations:

- The slowly perfused compartment was assumed to have a weighted average of the lipid and water content of the muscle and the bone.
- Rapidly perfused tissue was assumed to have the same lipid and water content as the brain, following the model of Krishnan et al. (2009).

The fraction unbound in the blood, fu_b was assumed to be the "free" percentage of the compound, as calculated using equations from (Feldman, 1972) and shown in Table 2.2.

The calculated values of the partition coefficients are found in Table 3.2. Note that we do not calculate the blood:blood partition coefficient using Equation (3.4), as it is clear that the ratio of the binding in blood to itself is equal to one since we have already accounted for binding of substrates.

Dynamic Adipose Partition Coefficients

For later iterations of the model, we allow the adipose tissue partition coefficient to adjust dynamically with the concentration of the compound in the venous compartment. This technique is novel in PBPK modeling–partition coefficients are typically assumed to be constant and reflect the partitioning for steady-state conditions. However, with the extremely high lipophilicity of the compounds and the daily dosing regimen, the adipose tissue may not achieve steady-state within a 24 hour period. By allowing the adipose PCs to vary with the venous concentration, we are able to more accurately reflect the sequestering/donating behavior of the adipose tissue (Rosenstreich et al., 1971). Equation (3.8) gives the form of the equation for varying the adipose PC for each compound:

$$P_{Adi}(t) = P_{Adi}^{est} - \frac{1}{2}P_{Adi}^{up} + \frac{P_{Adi}^{up}}{1 + \exp(\bar{x} - x(t))}, \tag{3.8}$$

where $P_{Adi}(t)$ is the adipose PC used in the PBPK equation at time t, P_{Adi}^{up} is an upper bound on estimations for the maximum adipose PC, P_{Adi}^{est} is the adipose PC predicted using the methods described previously, $\bar{x}$ is the steady-state concentration of the substrate in the venous compartment (as assumed from initial conditions), and $x(t)$ is the concentration of the substrate in the venous compartment at time t. When $x(t) = \bar{x}$, it is clear that $P_{Adi}(t) = P_{Adi}^{est}$. If $x(t) < \bar{x}$, the adipose PC will be less than predicted, reflecting donating of the substrate from the adipose compartment. As $x(t)$ gets much larger than the

Table 3.2: Parameters for adult humans and corresponding significance in the PBPK model. Parameters adapted from (Williams and Leggett, 1989), with the exception of volume percent of adipose tissue, which is given by Equation (3.3).

Compartment	**Symbol**	**Blood Flow %**	**Volume %**	**Partition Coefficient**		
				D	25(OH)D	1,25(OH)2D
Liver	L	7.0	2.40	5.5776	5.5759	5.5960
Kidney	K	18.0	0.44	3.3655	3.3645	3.3766
GI Tract	GI	16.0	1.76	7.0563	7.0541	7.0795
Skin	S	5.0	4.18	4.1779	4.1767	4.1917
Rapidly Perfused	RP	19.0	2.58	7.8555	7.8531	7.8814
Slowly Perfused	SP	25.25	54.33	4.5927	4.5913	4.6078
Adipose	Adi	6.75	26.64	31.3317	31.2786	31.9146
Spleen	Spl	3.0	0.23	3.4287	3.4277	3.4400
Arterial	A	100	1.86	1.0000	1.0000	1.0000
Venous	V	100	5.59	1.0000	1.0000	1.0000

steady-state concentration, $P_{Adi}(t)$ approaches P_{Adi}^{up}. This can be seen in Figure 3.5.

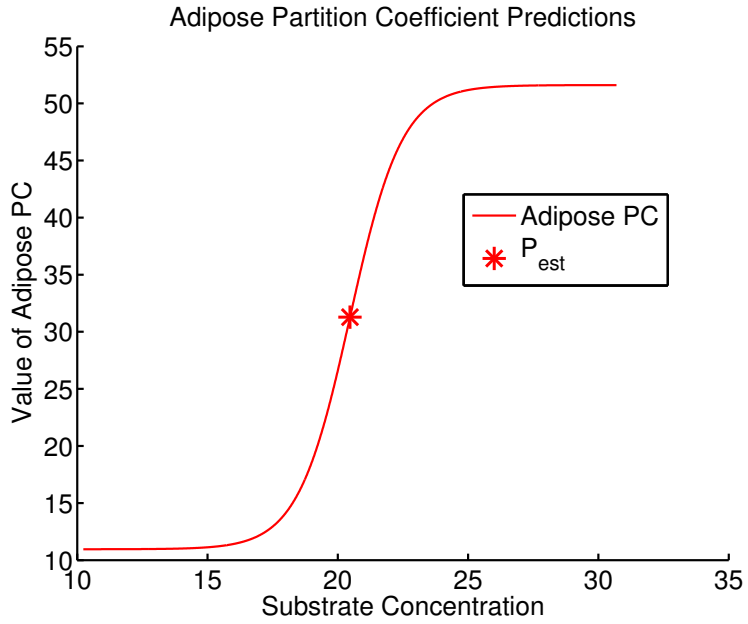


Figure 3.5: Example of $P_{Adi}(t)$ for 25(OH)D

3.2.5 Kinetic Parameters

For the purposes of this model, we consider four enzymes–CYPs 27A1, 27B1, 2R1, 24A1–to be involved in the conversion of vitamin D to its active form and subsequent inactivation. Each enzyme is assumed to follow standard one-site Michaelis-Menten kinetics (Sakaki et al., 2005), shown in Equation (3.9), where V is defined to be the rate of conversion of substrate to product:

$$V = \frac{V_{max}[S]}{K_m + [S]}. \tag{3.9}$$

Here, V_{max} is the maximum velocity of the enzymatic reaction, K_m is the concentration of the substrate at half of the maximum velocity, and $[S]$ is the concentration of the substrate. Note that at low concentrations of the substrate, Equation (3.9) can be approximated by

$$V = \frac{V_{max}}{K_m}[S]. \tag{3.10}$$

This comprises the linear aspect of the Michaelis-Menten curve, whereas higher substrate concentrations approximate the asymptotic portion of the Michaelis-Menten curve. For substrate concentrations observed by this study, we fall mostly in the linear portion of the curve. This will become apparent when performing sensitivity analysis, discussed in Section 4.2.

Ranges for relevant metabolic parameters were obtained from current literature. Whenever possible, research results from the same research group or with similar experimental protocols was used to minimize possible discrepancies due to variances in protocols. As an example, Tang et al. (2012) points out several differences in their experimental methods that resulted in much higher expression levels of CYP27B1 over previous studies.

A majority of literature values for V_{max} parameters are given in terms of moles of enzyme. To convert these values to usable units for modeling, the average amount of enzyme in the liver or kidney (depending on location) must be determined. We use the following conversion scheme (Equation (3.11)) to translate V_{max} $\left(\text{in units } \frac{\text{mol product}}{\text{time}\times\text{mol enzyme}}\right)$ parameters into usable values:

$$\frac{\text{mol product}}{\text{time}} = \frac{\text{mol product}}{\text{time} \times \text{mol enzyme}} \times \frac{\text{mol enzyme}}{\text{mg CYP}} \times \frac{\text{mg CYP}}{\text{g tissue}} \times \text{g tissue}. \tag{3.11}$$

After an extensive search of literature, we used estimates for conversion factors (aside

from the $\frac{\text{mol enzyme}}{\text{mg CYP}}$ term), found in Table 3.3. Literature reference values use the total

Table 3.3: Estimation of conversion factors for V_{max}. [1]mg/g tissue, [2]nmol/g protein. [a]Rajwade et al. (1975), [b]Clement et al. (2005), [c]Wilson et al. (2003), [d]Hrycay and Bandiera (2007), [e]Ghazarian et al. (1974).

	Mitochondrial		**Microsomal**	
Liver	protein frac[1]	% total CYP[2]	protein frac[1]	% total CYP[2]
	50[a]	0.31[b]	33[c]	0.3[d]
Kidney	protein frac[1]	% total CYP[2]	-	-
	64[a]	0.2[e]	-	-

amount of mitochondrial or microsomal CYPs per gram of tissue; to account for this, we add in the $\frac{\text{mol enzyme}}{\text{mg CYP}}$ term in Equation (3.11) to estimate the fraction that each enzyme comprises of the total amount of CYPs in the tissue. We have several estimates for the percentage distribution of CYP enzymes (Shimada et al., 1994). Metabolic studies in (Stubbs et al., 2010) suggest that after treatment, there is roughly three times as much CYP27B1 as CYP24A1 in the kidney, at least in patients with chronic kidney disease. Allowing for the notation $[enz] = \frac{\text{mol enzyme}}{\text{mg CYP}}$, we use the following initial estimates for the enzyme percentage of total CYPs (Equation (3.12)).

$$[27A1] = 1, \qquad [2R1] = 1, \qquad [27B1] = 3, \qquad [24A1] = 1. \tag{3.12}$$

Then V_{max} for a particular enzyme, as utilized in Equation (3.9), can be expressed as

$$V_{max} = V_{max,lit} \times [enz] \times (\text{protein frac}) \times (\%\ \text{total CYP}) \times \text{g tissue}, \tag{3.13}$$

where $V_{max,lit}$ is the literature-obtained value, $[enz]$ is as described in Equation (3.12), and the remaining parameters as described in Table 3.3 with the appropriate choice of tissue and enzyme location.

Regulatory Considerations

As discussed in Section 2.2.5, the concentration of 1,25(OH)2D can influence the amount of CYP27B1 and CYP24A1 available to the cascade. The initial PBPK model assumes that the amount of CYPs 27B1 and 24A1 do not change significantly; however, to improve the fit for later time points, the final PBPK model allows for these amounts to change in relation to the concentration of 1,25(OH)2D in the kidney. Equations (3.14) and (3.15) give the general forms of the equations that we use to describe the change in $[27B1]$ and $[24A1]$, respectively, based on the amount of 1,25(OH)2D in the kidney:

$$[27B1]_t = [27B1]_0 - [27B1]_0 \times \left(\frac{2}{1 + \exp(2\alpha_1(x(0) - x(t)))} - 1 \right) \tag{3.14}$$

$$[24A1]_t = [24A1]_0 + [24A1]_0 \times \left(\frac{2}{1 + \exp(2\alpha_2(x(0) - x(t)))} - 1 \right). \tag{3.15}$$

Here, $[27B1]$ and $[24A1]$ are the percentage estimates at time t, where $t = 0$ is the initial estimate, for CYPs 27B1 and 24A1, respectively, and x is the concentration of 1,25(OH)2D in the kidney at time t, where $t = 0$ is the initial concentration in the kidney. The parameters α_1 and α_2 control the "steepness" of the approach to either asymptote on the sigmoidal curve, and are optimized for in this model. This is physiologically represented by the speed at which concentrations of 1,25(OH)2D influence the transcriptional activation or deactivation of enzymes. Figure 3.6 demonstrates the shape of these sigmoidal curves.

Several different functions could have been chosen to represent this biological phe-

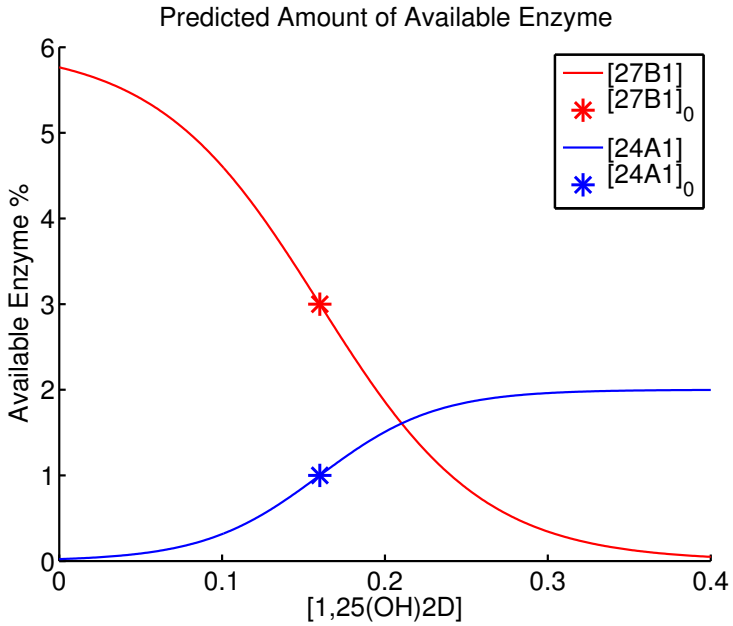


Figure 3.6: Sigmoidal curves representing Equations (3.14) and (3.15)

nomenon; we chose to use a sigmoidal function for several reasons.

1. Sigmoidal functions are saturating and nonlinear. Biologically-speaking, we expect the change in transcription levels to be tightly bound and not exceed specific upper and lower bounds. For the purposes of this function, there is always some amount of either CYP27B1 or CYP24A1 produced, so 0% is a relevant lower bound. To simplify the model, we allowed for the upper bound to be twice the initial percentage of the enzyme. This allows for the upper and lower bounds on the sigmoidal curve to be equidistant from the initial percentage. Throughout our experimental runs, these bounds are more than sufficient as the varying percent does not come significantly close to either asymptote.

2. Sigmoidal functions are often used in enzyme kinetics to describe enzymes with positive cooperative binding of substrates. In this case, 1,25(OH)2D acts to increase

the total amount of substrate bound to CYP24A1 by increasing the amount of CYP24A1 available for binding. This increases the apparent V_{max} of the reactions involving CYP24A1. Similarly, with the addition of the phase shift in the sigmoidal function, 1,25(OH)2D acts to decrease the total amount of substrate bound to CYP27B1 by decreasing the amount of CYP27B1 available (and thus the apparent V_{max} of the CYP27B1 reactions).

3. The choice of a simple adaptation of the logistic function $\frac{1}{1+e^{-t}}$ was chosen over other sigmoidal curves, such as the Hill or Gompertz functions, because it is symmetric about the inflection point (Ratkowsky and Giles, 1990).

3.2.6 Absorption and Excretion

Vitamin D has two methods of entry into the human body. The first is synthesis in the skin, as discussed in Section 2.2.1, and the second is through oral consumption. We control for dermal synthesis of vitamin D by utilizing a data set from an experiment conducted when sun exposure is minimal (e.g., during an English winter). The only route of entry is, therefore, ingestion. Several studies have attempted to quantify the rate at which vitamin D is absorbed into the body either through synthesis or ingestion, mostly in rats. We chose to focus on a study of *in vivo* absorption in rats (Hollander et al., 1978), in which a linear relationship between the absorption of vitamin D and its concentration in the GI tract was established. Accordingly, we define the rate of absorption, Abs, to be a function of the concentration of vitamin D in the GI tract, shown in Equation (3.16):

$$Abs = 3.2 \times b + \frac{1}{6} \times A_{Lumen}, \tag{3.16}$$

where b is the "free" percentage of the dose and A_{Lumen} is the amount of vitamin D present in the GI lumen. Participants ingest one of three dosing amounts of vitamin D on a daily basis, which results in a pulsing effect on the amount in the GI lumen.

A variety of studies have been done on the excretion of vitamin D, the most cited of which is (Avioli et al., 1967). This study tracked the disappearance of radio-labeled, intravenously administered vitamin D from the plasma. At the time of this study, metabolites of vitamin D hadn't been clearly identified, so the total amount of label excreted in urine and feces included label contained in metabolites of vitamin D. However, we were able to utilize the data contained in this study to get an approximation of the amount of vitamin D excreted in the urine and feces per hour. This amount is so small that there was no significant change in model output with the inclusion of excretion terms. To increase the predictive power of our model by reducing the number of unknown parameters, we therefore chose to leave out excretion of vitamin D through the urine or feces.

3.3 Mathematical Model

As previously stated, each enzyme is assumed to follow standard one-site Michaelis-Menten kinetics. Also detailed were changes to the apparent V_{max} by adjusting the available levels of enzymes for reactions involving CYPs 27B1 and 24A1. We can group the metabolic reactions based on parent compound and enzyme location in the following manner. Let Met_j^i be the metabolism occurring in compartment j with parent compound i where j is either the liver or the kidney and i is in the set {D; 25; 1,25} where we use the abbreviation "25" for 25(OH)D and "1,25" for 1,25(OH)2D. Superscripts on the kinetic parameters indicate the enzyme responsible. C_j^i indicates the concentration of the compound i in

compartment j, and is defined as

$$C_j^i = \frac{A_j^i}{V_j P_j^i} \tag{3.17}$$

for A amount, V volume, and P the tissue:blood partition coefficient.

Liver Metabolism:

$$Met_L^D = -\frac{V_{max}^{27A1} C_L^D}{K_m^{27A1} + C_L^D} - \frac{V_{max}^{2R1} C_L^D}{K_m^{2R1} + C_L^D} \tag{3.18}$$

$$Met_L^{25} = \frac{V_{max}^{27A1} C_L^D}{K_m^{27A1} + C_L^D} + \frac{V_{max}^{2R1} C_L^D}{K_m^{2R1} + C_L^D} - \frac{V_{max}^{27A1b} C_L^{25}}{K_m^{27A1b} + C_L^{25}} \tag{3.19}$$

$$Met_L^{1,25} = \frac{V_{max}^{27A1b} C_L^{25}}{K_m^{27A1b} + C_L^{25}} \tag{3.20}$$

Kidney Metabolism:

$$Met_K^D = 0 \tag{3.21}$$

$$Met_K^{25} = -\frac{V_{max}^{27B1} C_K^{25}}{K_m^{27B1} + C_K^{25}} - \frac{V_{max}^{24A1b} C_K^{25}}{K_m^{24A1b} + C_K^{25}} \tag{3.22}$$

$$Met_K^{1,25} = \frac{V_{max}^{27B1} C_K^{25}}{K_m^{27B1} + C_K^{25}} - \frac{V_{max}^{24A1} C_K^{1,25}}{K_m^{24A1} + C_K^{1,25}} \tag{3.23}$$

The PBPK model for vitamin D can be described by

$$\dot{\mathbf{x}} = \mathbf{f}(t, \mathbf{x}, \mathbf{q}) \tag{3.24}$$

where $\mathbf{x} \in \mathbb{R}^{31}$ is the states, t is time, and $\mathbf{q}$ is model parameters. The total equations used in $\mathbf{f}$ are as follows:

- $\mathbf{x}$ has 31 dimensions, where
  - 30 states are given by Equations (3.25) to (3.54);

  - GI lumen given by Equation (3.16).

- 6 kinetic equations given by Equations (3.18) to (3.23);
- 30 partition coefficient equations of the form Equations (3.4) and (3.5);
  - The adipose partition coefficients are modified using Equation (3.8).
- 2 equations describing enzyme concentrations (Equations (3.14) and (3.15)).

For the following equations, grouped by metabolite for convenience, we define the hepatic blood flow Q_{hep} as $Q_{hep} = Q_L + Q_{GI} + Q_{Spl}$, and $\Lambda = \{K, S, RP, SP, Adi\}$ defines the compartments used in calculating the rate of change for the venous compartment. CA and CV are the concentrations in the arterial and venous compartments, respectively.

Vitamin D:

$$\text{Liver: } \frac{dA_L^D}{dt} = Q_L CA^D + Q_{GI} C_{GI}^D + Q_{Spl} C_{Spl}^D - Q_{hep} C_L^D + Met_L^D \tag{3.25}$$

$$\text{Kidney: } \frac{dA_K^D}{dt} = Q_K (CA^D - C_K^D) \tag{3.26}$$

$$\text{GI Tract: } \frac{dA_{GI}^D}{dt} = Q_{GI} (CA^D - C_{GI}^D) + Abs \tag{3.27}$$

$$\text{Skin: } \frac{dA_S^D}{dt} = Q_S (CA^D - C_S^D) \tag{3.28}$$

$$\text{Rap. Per.: } \frac{dA_{RP}^D}{dt} = Q_{RP} (CA^D - C_{RP}^D) \tag{3.29}$$

$$\text{Slow. Per.: } \frac{dA_{SP}^D}{dt} = Q_{SP} (CA^D - C_{SP}^D) \tag{3.30}$$

$$\text{Adipose: } \frac{dA_{Adi}^D}{dt} = Q_{Adi} (CA^D - C_{Adi}^D) \tag{3.31}$$

$$\text{Spleen: } \frac{dA^D_{Spl}}{dt} = Q_{Spl}(CA^D - C^D_{Spl}) \tag{3.32}$$

$$\text{Arterial: } \frac{dAA^D}{dt} = Q_{Art}(CV^D - CA^D) \tag{3.33}$$

$$\text{Venous: } \frac{dAV^D}{dt} = \sum_{Z\in\Lambda} Q_Z C^D_Z + Q_{hep}C^D_L - Q_{Ven}CV^D \tag{3.34}$$

25(OH)D:

$$\text{Liver: } \frac{dA^{25}_L}{dt} = Q_L CA^{25} + Q_{GI}C^{25}_{GI} + Q_{Spl}C^{25}_{Spl} - Q_{hep}C^{25}_L + Met^{25}_L \tag{3.35}$$

$$\text{Kidney: } \frac{dA^{25}_K}{dt} = Q_K(CA^{25} - C^{25}_K) + Met^{25}_K \tag{3.36}$$

$$\text{GI Tract: } \frac{dA^{25}_{GI}}{dt} = Q_{GI}(CA^{25} - C^{25}_{GI}) \tag{3.37}$$

$$\text{Skin: } \frac{dA^{25}_S}{dt} = Q_S(CA^{25} - C^{25}_S) \tag{3.38}$$

$$\text{Rap. Per.: } \frac{dA^{25}_{RP}}{dt} = Q_{RP}(CA^{25} - C^{25}_{RP}) \tag{3.39}$$

$$\text{Slow. Per.: } \frac{dA^{25}_{SP}}{dt} = Q_{SP}(CA^{25} - C^{25}_{SP}) \tag{3.40}$$

$$\text{Adipose: } \frac{dA^{25}_{Adi}}{dt} = Q_{Adi}(CA^{25} - C^{25}_{Adi}) \tag{3.41}$$

$$\text{Spleen: } \frac{dA^{25}_{Spl}}{dt} = Q_{Spl}(CA^{25} - C^{25}_{Spl}) \tag{3.42}$$

$$\text{Arterial: } \frac{dAA^{25}}{dt} = Q_{Art}(CV^{25} - CA^{25}) \tag{3.43}$$

$$\text{Venous: } \frac{dAV^{25}}{dt} = \sum_{Z\in\Lambda} Q_Z C^{25}_Z + Q_{hep}C^{25}_L - Q_{Ven}CV^{25} \tag{3.44}$$

1,25(OH)2D:

$$\text{Liver: } \frac{dA_L^{1,25}}{dt} = Q_L CA^{1,25} + Q_{GI} C_{GI}^{1,25} + Q_{Spl} C_{Spl}^{1,25} - Q_{hep} C_L^{1,25} + Met_L^{1,25} \tag{3.45}$$

$$\text{Kidney: } \frac{dA_K^{1,25}}{dt} = Q_K (CA^{1,25} - C_K^{1,25}) + Met_K^{1,25} \tag{3.46}$$

$$\text{GI Tract: } \frac{dA_{GI}^{1,25}}{dt} = Q_{GI} (CA^{1,25} - C_{GI}^{1,25}) \tag{3.47}$$

$$\text{Skin: } \frac{dA_S^{1,25}}{dt} = Q_S (CA^{1,25} - C_S^{1,25}) \tag{3.48}$$

$$\text{Rap. Per.: } \frac{dA_{RP}^{1,25}}{dt} = Q_{RP} (CA^{1,25} - C_{RP}^{1,25}) \tag{3.49}$$

$$\text{Slow. Per.: } \frac{dA_{SP}^{1,25}}{dt} = Q_{SP} (CA^{1,25} - C_{SP}^{1,25}) \tag{3.50}$$

$$\text{Adipose: } \frac{dA_{Adi}^{1,25}}{dt} = Q_{Adi} (CA^{1,25} - C_{Adi}^{1,25}) \tag{3.51}$$

$$\text{Spleen: } \frac{dA_{Spl}^{1,25}}{dt} = Q_{Spl} (CA^{1,25} - C_{Spl}^{1,25}) \tag{3.52}$$

$$\text{Arterial: } \frac{dAA^{1,25}}{dt} = Q_{Art} (CV^{1,25} - CA^{1,25}) \tag{3.53}$$

$$\text{Venous: } \frac{dAV^{1,25}}{dt} = \sum_{Z \in \Lambda} Q_Z C_Z^{1,25} + Q_{hep} C_L^{1,25} - Q_{Ven} CV^{1,25} \tag{3.54}$$

3.4 Existence and Uniqueness

We want to guarantee that a solution to $\mathbf{f}$ exists and is unique. Showing existence here will allow us to justify sensitivity equations later. We will first show that the basic PBPK model described in Section 3.2.1 satisfies existence and uniqueness criteria, after which we

will show that the adaptions to the model (discussed in Section 3.2.4 and Section 3.2.5) maintain these criteria. We will follow the outline of the proof in Manning et al. (2010) to show existence and uniqueness.

3.4.1 Basic Model

Let $\mathbf{f} = [f_1, f_2, \ldots, f_n]'$ be such that $\mathbf{f}(\mathbf{x}) : U \to \mathbb{R}^n$ for U an open subset of $\mathbb{R}^n$. Consider the autonomous system of nonlinear differential equations

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) \tag{3.55}$$

with $\mathbf{x}(t_0) = \mathbf{x}_0 \in \mathbb{R}^n$ and $t_0 \in \mathbb{R}$. To determine if $\mathbf{f}$ is locally Lipschitz, we can examine each component, f_i. In this model for $i = 1, 2, \ldots, n$ with $n = 30$, each f_i has the form

$$f_i(\mathbf{x}) = A_i + \sum_{j=1}^{n} B_{i_j} x_j H(x_j) - \sum_{k=1}^{n} C_{i_k} x_k H(x_k) + \sum_{l=1}^{n} D_{i_l} \frac{x_l H(x_l)}{x_l H(x_l) + K_l} \tag{3.56}$$

where $A_i, B_{i_j}, C_{i_k}, D_{i_k}$, and K_l are nonnegative constants with the following definitions:

- x_j is the concentration in the jth compartment and is the jth component of $\mathbf{x}$;
- H is the Heaviside function given by

$$H(x_j) = \begin{cases} 1 & x_j \geq 0 \\ 0 & x_j < 0 \end{cases} . \tag{3.57}$$

  The Heaviside function is used to guarantee that each component of $\mathbf{x}$ is nonnegative. This is reasonable since the components of $\mathbf{x}$ represent physiological concentrations, which require non-negativity to be plausible.

- B_{i_j} corresponds to flow into compartment i from compartment j;

- C_{i_k} corresponds to flow out of compartment i into compartment k, incorporating the i:k partition coefficient;
- D_{i_l} represents the maximum velocity of the enzymatic reaction, V_{max}, as the compound moves from compartment i to l; and
- K_l is the concentration of x_l when $\frac{1}{2}V_{max}$ is achieved.

Theorem 3.4.1. *The autonomous system of nonlinear differential equations, as described by Equation* (3.55)*, is locally Lipschitz.*

Proof. We will first show that each f_i, as described by Equation (3.56) is locally Lipschitz. Let

$$N_\varepsilon(\mathbf{x}_0) = \{\mathbf{x} \in \mathbb{R}^n \Big| \, \|\mathbf{x} - \mathbf{x}_0\| < \varepsilon\}$$

where $\varepsilon > 0$ and the norm $\|\cdot\|$ is any equivalent norm in $\mathbb{R}^n$. Let $\vec{\phi}$ and $\vec{\psi}$ be vectors in $N_\varepsilon(\mathbf{x}_0)$. Now consider the absolute value of the difference in f_i evaluated at $\vec{\phi}$ and $\vec{\psi}$.

$$\begin{aligned}
\left|f_i(\vec{\phi}) - f_i(\vec{\psi})\right| &= \left| \left\{ A_i + \sum_{j=1}^{n} B_{i_j}\phi_j H(\phi_j) - \sum_{k=1}^{n} C_{i_k}\phi_k H(\phi_k) + \sum_{l=1}^{n} D_{i_l} \frac{\phi_l H(\phi_l)}{\phi_l H(\phi_l) + K_l} \right\} \right. \\
&\qquad \left. - \left\{ A_i + \sum_{j=1}^{n} B_{i_j}\psi_j H(\psi_j) - \sum_{k=1}^{n} C_{i_k}\psi_k H(\psi_k) + \sum_{l=1}^{n} D_{i_l} \frac{\psi_l H(\psi_l)}{\psi_l H(\psi_l) + K_l} \right\} \right| \\
&= \left| \sum_{j=1}^{n} B_{i_j}\phi_j H(\phi_j) - \sum_{k=1}^{n} C_{i_k}\phi_k H(\phi_k) + \sum_{l=1}^{n} D_{i_l} \frac{\phi_l H(\phi_l)}{\phi_l H(\phi_l) + K_l} \right. \\
&\qquad \left. - \sum_{j=1}^{n} B_{i_j}\psi_j H(\psi_j) + \sum_{k=1}^{n} C_{i_k}\psi_k H(\psi_k) - \sum_{l=1}^{n} D_{i_l} \frac{\psi_l H(\psi_l)}{\psi_l H(\psi_l) + K_l} \right| \\
&\leq \sum_{j=1}^{n} \left| B_{i_j}[\phi_j H(\phi_j) - \psi_j H(\psi_j)] \right| + \sum_{k=1}^{n} \left| C_{i_k}[\psi_k H(\psi_k) - \phi_k H(\phi_k)] \right|
\end{aligned}$$

$$+\sum_{l=1}^{n}\left|D_{i_l}\left[\frac{\phi_l H(\phi_l)}{\phi_l H(\phi_l)+K_l}-\frac{\psi_l H(\psi_l)}{\psi_l H(\psi_l)+K_l}\right]\right|$$

$$=\sum_{j=1}^{n}\left|B_{i_j}[\phi_j H(\phi_j)-\psi_j H(\psi_j)]\right|+\sum_{k=1}^{n}\left|C_{i_k}[\psi_k H(\psi_k)-\phi_k H(\phi_k)]\right|$$

$$+\sum_{l=1}^{n}\left|D_{i_l}\left[\frac{K_l\Big(\phi_l H(\phi_l)-\psi_l H(\psi_l)\Big)}{\phi_l H(\phi_l)\psi_l H(\psi_l)+K_l\Big(\phi_l H(\phi_l)+\psi_l H(\psi_l)\Big)+K_l^2}\right]\right|$$

$$\leq\sum_{j=1}^{n}\left|B_{i_j}[\phi_j H(\phi_j)-\psi_j H(\psi_j)]\right|+\sum_{k=1}^{n}\left|C_{i_k}[\psi_k H(\psi_k)-\phi_k H(\phi_k)]\right|$$

$$+\sum_{l=1}^{n}\left|D_{i_l}\left[\frac{K_l\Big(\phi_l H(\phi_l)-\psi_l H(\psi_l)\Big)}{K_l^2}\right]\right|$$

$$\leq\sum_{j=1}^{n}\left|B_{i_j}\right|\left|\phi_j H(\phi_j)-\psi_j H(\psi_j)\right|+\sum_{k=1}^{n}\left|C_{i_k}\right|\left|\psi_k H(\psi_k)-\phi_k H(\phi_k)\right|$$

$$+\sum_{l=1}^{n}\left|\frac{D_{i_l}}{K_l}\right|\left|\phi_l H(\phi_l)-\psi_l H(\psi_l)\right|. \tag{3.58}$$

Note that $|B_{i_j}|, |C_{i_k}|$, and $\left|\frac{D_{i_l}}{K_l}\right|$ are bounded and K_l, in particular, is the concentration at the half saturation point of the enzyme. To show Lipschitz continuity, we need to evaluate various cases for ϕ_k and ψ_k to show that $\left|\phi_j H(\phi_j)-\psi_j H(\psi_j)\right|$ is bounded.

1. If $\phi_j, \psi_j \geq 0$, then

$$\left|\phi_j H(\phi_j)-\psi_j H(\psi_j)\right| = \left|\phi_j-\psi_j\right|$$

$$\leq \|\phi-\psi\|.$$

2. If $\phi_j, \psi_j < 0$, then

$$\begin{aligned}\left|\phi_j H(\phi_j) - \psi_j H(\psi_j)\right| &= 0 \\ &\leq \|\phi - \psi\|.\end{aligned}$$

3. If $\phi_j \geq 0$ and $\psi_j < 0$, then coupled with the definition of absolute value, we obtain the following:

$$\begin{aligned}\left|\phi_j H(\phi_j) - \psi_j H(\psi_j)\right| &= \left|\phi_j\right| \\ &< \left|\left|\phi_j\right| + \left|\psi_j\right|\right| \\ &= \left|\phi_j - \psi_j\right| \quad \text{since } \psi_j < 0 \\ &\leq \|\phi - \psi\|.\end{aligned}$$

The same holds if $\psi_j \geq 0$ and $\phi_j < 0$.

Thus, for any ϕ_j and any ψ_j,

$$\left|\phi_j H(\phi_j) - \psi_j H(\psi_j)\right| \leq \|\phi - \psi\|.$$

This argument holds with the changed subscript as well.

Thus, we have shown that each term in Equation (3.58) is locally Lipschitz, so each f_i is locally Lipschitz. Therefore, existence and uniqueness of a solution to the system $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$ with $\mathbf{x}(t_0) = \mathbf{x}_0$ is guaranteed by Theorem 3.4.2.

□

Theorem 3.4.2 (Existence and Uniqueness of Solutions of Ordinary Differential Equa-

tions). *Let $U \subset \mathbb{R}^n$ be an open set and let $\mathbf{f} : U \to \mathbb{R}^n$ be a (local) Lipschitz function. Let $\mathbf{x}_0 \in U$ and $t_0 \in \mathbb{R}$. Then there exists an $\alpha > 0$ and a solution $\mathbf{x}(t)$ of $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$ defined for $t \in (t_0 - \alpha, t_0 + \alpha)$ such that $\mathbf{x}(t_0) = \mathbf{x}_0$. Moreover, if $\mathbf{y}(t)$ is another solution with $\mathbf{y}(t_0) = \mathbf{x}_0$, then $\mathbf{x}(t) = \mathbf{y}(t)$ on their common interval of definition about t_0. (Robinson, 1999)*

3.4.2 Dynamic Adipose Partition Coefficients

In the first adaption to the basic model, we allow the partition coefficients for the adipose tissue compartment to vary in accordance with the concentration of the compound in the venous compartment. This relation is given by

$$P_i(x) = P_i^{est} - \frac{1}{2}P_i^{up} + \frac{P_i^{up}}{1 + \exp(\bar{x} - xH(x))}, \tag{3.59}$$

where $P_i(x)$ is the the adipose partition coefficient relating to the concentration x of the compound in the venous compartment for $i \in \{D; 25; 1, 25\}$; P_i^{est} is the adipose PC of compound i estimated using the methods of (Poulin and Theil, 2002); P_i^{up} is an estimate for the upper bound of the adipose PC, which can be no higher than 130% of P_i^{est} by model design; and $\bar{x}$ is the steady-state amount of compound i in the venous compartment, as initialized by the model. For completeness, we allow for the possibility of a negative concentration, x, through the incorporation of a Heaviside function term $H(x)$, as before.

The partition coefficients are incorporated in the C_{i_k} term in Equation (3.56). We have previously shown that if C_{i_k} is constant and bounded by assumption, then each f_i is locally Lipschitz. If a term in C_{i_k} varies, then we need to show that the term is still bounded to ensure that f_i is also bounded.

Since each C_{i_k} is composed of the ratio of a flow rate Q_i and partition coefficient P_i^{est},

we must show that $\frac{1}{P_i(x)}$ is bounded, or equivalently, that $P_i(x)$ is bounded away from zero. We will show this by establishing that $P_i(x)$ is a strictly increasing function, and then by establishing its positivity. We will conclude this proof by establishing upper and lower bounds on $P_i(x)$, and consequently, on $\frac{1}{P_i(x)}$.

Lemma 3.4.3. *Let* $\dot{\mathbf{x}} = \hat{\mathbf{f}}(\mathbf{x})$ *where* $\hat{\mathbf{f}}$ *is* $\mathbf{f}$ *as described in Theorem 3.4.1 with modified adipose tissue partition coefficients as described by Equation* (3.59)*. Then this system is locally Lipschitz.*

Proof. To show that each C_{i_k} is bounded, we will determine bounds on $P_i(x)$. We will first establish that $P_i(x)$ is a strictly increasing function. Let $DP_i(x)$ be the derivative of $P_i(x)$ with respect to x. Then

$$\begin{aligned}
DP_i(x) &= \frac{d}{dx}\left[P_i^{est} - \frac{1}{2}P_i^{up} + \frac{P_i^{up}}{1+\exp(\bar{x} - xH(x))}\right] \\
&= P_i^{up}\frac{d}{dx}\left[\frac{1}{1+\exp(\bar{x} - xH(x))}\right] \\
&= -\frac{P_i^{up}\exp(\bar{x} - xH(x))}{[1+\exp(\bar{x} - xH(x))]^2}\frac{d}{dx}[\bar{x} - xH(x)] \\
&= -\frac{P_i^{up}\exp(\bar{x} - xH(x))}{[1+\exp(\bar{x} - xH(x))]^2}[-x\delta(x) - H(x)] \\
&= \frac{P_i^{up}\exp(\bar{x} - xH(x))}{[1+\exp(\bar{x} - xH(x))]^2}[x\delta(x) + H(x)], \qquad (3.60)
\end{aligned}$$

for $\delta(x)$ the Dirac delta function. For $x \leq 0$, $DP_i(x) = 0$. For $x > 0$, $DP_i(x) > 0$ since $P_i^{up} > 0$ and $\exp(\bar{x} - xH(x)) > 0$; thus Equation (3.60) is strictly positive in the physiological region of interest. This means that Equation (3.59) is a strictly increasing function.

To show that $P_i(x)$ is a positive function, we will establish a lower bound. Consider $\lim_{x\to-\infty} P_i(x)$. Note that $x < 0$ is not physiologically reasonable, but since $P_i(x)$ is increasing along its entire domain, we can utilize the mathematical range of the function to establish a lower bound before considering the physiological range of the function.

$$\begin{aligned}
\lim_{x\to-\infty} P_i(x) = & \lim_{x\to-\infty}\left[P_i^{est} - \frac{1}{2}P_i^{up} + \frac{P_i^{up}}{1+\exp(\bar{x} - xH(x))}\right] \\
= & \ P_i^{est} - \frac{1}{2}P_i^{up} + P_i^{up}\lim_{x\to-\infty}\frac{1}{1+\exp(\bar{x} - xH(x))} \\
\geq & \ P_i^{est} - \frac{1}{2}P_i^{up} \\
\geq & \ P_i^{est} - .65P_i^{est} \\
= & \ .35\ P_i^{est} > 0, \qquad (3.61)
\end{aligned}$$

since by model design, $P_i^{up} \leq 1.3\ P_i^{est}$ and $P_i^{est} > 0$. Thus, the mathematical minimum of $P_i(x)$ is greater than zero, so $\frac{1}{P_i(x)}$ is bounded below by $\frac{1}{.35\ P_i(x)}$.

Furthermore, we can establish an upper bound on $P_i(x)$ in the same fashion.

$$\begin{aligned}
\lim_{x\to\infty} P_i(x) = & \lim_{x\to\infty}\left[P_i^{est} - \frac{1}{2}P_i^{up} + \frac{P_i^{up}}{1+\exp(\bar{x} - xH(x))}\right] \\
= & \ P_i^{est} - \frac{1}{2}P_i^{up} + P_i^{up}\lim_{x\to\infty}\frac{1}{1+\exp(\bar{x} - xH(x))} \\
= & \ P_i^{est} + \frac{1}{2}P_i^{up} \\
\leq & \ P_i^{est}(1+.65) = 1.65\ P_i^{est}. \qquad (3.62)
\end{aligned}$$

Combining the results from Equations (3.61) and (3.62) yields the following bounds for the adipose tissue partition coefficient for compound i:

$$0 < \frac{1}{1.65\ P_i^{est}} \leq \frac{1}{P_i(x)} \leq \frac{1}{.35\ P_i^{est}}. \tag{3.63}$$

Thus C_{i_k} is bounded, even with variable partition coefficients.

□

We have shown that even with the addition of variable partition coefficients, each f_i, as defined in Equation (3.56), is locally Lipschitz, thus the overall model satisfies existence and uniqueness criteria.

3.4.3 Transcriptional Regulation of Enzyme Concentrations

A second modification to the base model was to allow transcriptional regulation of the availability of CYPs 27B1 and 24A1, which is reflected in the associated V_{max} parameters. As discussed in Section 3.2.5, we utilized a sigmoidal function to describe the percent of total CYPs composed of either CYP27B1 or CYP24A1. Recall the following equations:

$$[27B1]_t = [27B1]_0 - [27B1]_0 \times \left(\frac{2}{1 + \exp(2\alpha_1(F_0 - F_t))} - 1 \right) \tag{3.64}$$

$$[24A1]_t = [24A1]_0 + [24A1]_0 \times \left(\frac{2}{1 + \exp(2\alpha_2(F_0 - F_t))} - 1 \right). \tag{3.65}$$

Lemma 3.4.4. *Let $\dot{\mathbf{x}} = \mathbf{g}(\mathbf{x})$ where $\mathbf{g}$ is $\mathbf{f}$ as described in Theorem 3.4.1 and has varying total percentages of CYPs 27B1 and 24A1 as described in Equations* (3.64) *and* (3.65)*. Then this system is locally Lipschitz.*

Proof. We will follow the same techniques as in the proof for Lemma 3.4.3.

Consider first Equation (3.64). Physiologically, we expect that as the concentration of 1,25(OH)2D increases (corresponding to an increase in F_t), transcriptional regulation of enzyme production will cause less CYP27B1 to be produced, which will, in turn, decrease the amount of 25(OH)D metabolized into 1,25(OH)2D. We expect that Equation (3.64) will be strictly decreasing as F_t increases. Let $D[27B1]_t$ be the derivative of $[27B1]_t$ with respect to F_t. Then

$$\begin{aligned} D[27B1]_t = & \frac{d}{dF_t}\left[[27B1]_0 - [27B1]_0 \times \left(\frac{2}{1+\exp(2\alpha_1(F_0 - F_t))} - 1\right)\right] \\ & = -\frac{4\alpha_1[27B1]_0 \exp(2\alpha_1(F_0 - F_t))}{[1+\exp(2\alpha_1(F_0 - F_t))]^2}. \end{aligned}$$

All terms in the fraction are positive, so $D[27B1]_t < 0$, which means $[27B1]_t$ is strictly decreasing as F_t increases. The following limits are clear:

$$\lim_{F_t \to -\infty} = 2[27B1]_0 \qquad \text{and} \qquad \lim_{F_t \to \infty} = 0, \tag{3.66}$$

so $0 < [27B1]_t < 2[27B1]_0$.

We can show a similar proof to illustrate that Equation (3.65) is strictly increasing and that $0 < [24A1]_t < 2[24A1]_t$.

The $[27B1]_t$ and $[24A1]_t$ come into play when calculating V_{max} for each enzyme, which is represented by the term D_{i_l} in Equation (3.56). For the proof of Theorem 3.4.1 to hold, we require that D_{i_l} remains nonnegative and bounded, which remains true with Equations (3.64) and (3.65).

□

Chapter 4: Model Calibration and Subset Selection

The first goal after the construction of a model is calibration. Calibration occurs when the model parameters are fit to obtained data such that the resulting model predictions match the data as near as possible. This process is also known as solving an inverse problem where calibration of the model is based on the expected outcomes. Inverse problems are a staple part of PBPK modeling, as the data obtained from experiments rarely contains all relevant parameters for the PBPK model itself. Inverse problems also show up in the case of engineering and other physically-related problems. The interested reader is referred to (Banks and Tran, 2009) for an overview of inverse problems and methodology.

Inverse problems are typically difficult to solve because of common nonlinearities and ill-posedness issues. States and parameters may not be directly measurable or even observable and available data may be sparse or subject to measuring errors. In the case of the vitamin D PBPK model, the data available is sparse, with measurements occurring approximately every seven days and only in one compartment. Furthermore, the available data is on 25(OH)D, yielding little information about the metabolic processes that occur

before or after that point.

It is haphazard to begin parameter estimation without determining which parameters are sensitive–meaning that changes in the parameter value will change the behavior of the model–and of those parameters which are sensitive, which are also identifiable by the model. Even with adherence to sensitivity and identifiability techniques, there is still a degree of problem-specific expertise that must be applied; a parameter may be deemed insensitive in relation to other parameters but is still essential in describing model behavior when considered in a different manner. An example of this will be discussed with the case of the regulatory parameters affecting availability of pools of enzyme.

This chapter will first discuss several methodologies for sensitivity and identifiability analysis used to identify vitamin D model parameters available for estimation. After an introduction of techniques, applications of these techniques to the vitamin D PBPK model will be demonstrated.

4.1 Identifiability and Estimability

The differences between identifiability and estimability can be subtle. Identifiability, in its most basic sense, is often considered to be a form of *a-priori* analysis, relating only to the model structure. In contrast, estimability relies on experimental data, and is thus considered a form of *a-posteriori* analysis. A comparison of identifiability and estimability properties are summarized from (McLean et al., 2012) and are found in Table 4.1.

Identifiability analysis can indicate problems in model structure, whereas estimability can help determine if experimental data (whether in-hand or proposed) allows for unique estimation of parameters. If parameters are found to be unidentifiable, and if the identifiability of the parameters is important to model predictions, then the model must be

Table 4.1: A comparison of identifiability and estimability (summarized from Table 1 in McLean et al. (2012))

Concept	**Identifiability**	**Estimability**
Question	Can different parameter values lead to the same model behaviors?	Can all parameters be uniquely estimated using available experimental data?
Alternative Names	Structural Identifiability Theoretical Identifiability Qualitative Identifiability	Practical Identifiability Quantitative Identifiability
Information Used	Model equations Types of measurements	Model equations Types of measurements Experimental settings Initial parameter guesses
Mathematical Techniques	Linearization Taylor Series Expansion	Fisher Information Matrix Graphical or Visual Inspection Correlation matrix
Model Complexity	Difficult for more than 10 parameters and states	Can assess in models with more than 50 parameters and states

simplified or additional output variables (such as the concentration of another metabolite in the case of the vitamin D model) must be measured to achieve identifiability. As discussed in McLean et al. (2012), there are several ways to compensate for unidentifiable parameters:

1. Lump parameters that always appear together, i.e. if k_1 and k_2 always show up as a product k_1k_2, consider the parameter $k_3 = k_1k_2$ instead;

2. Remove terms from the model equation that are expected to have little influence on the model predictions; and

3. Fix some parameters at reasonable values to reduce the number of parameters subject to estimation.

In contrast, compensating for inestimable parameters may be harder, as inestimability may come from a lack of sensitivity of the model to changes in the parameter value or changes in one parameter set may be correlated with changes in another parameter set. To assess estimability of parameters, both the sensitivity of model predictions to each parameter as well as correlation between parameters must be considered. Again, we emphasize that these are local results, as different parameter sets may show varying degrees of correlation and model sensitivity.

4.1.1 (Structural) Identifiability

If parameters in a model are troublesome to identify, it could be due to either the model itself not having an appropriate structure or a lack of experimental evidence. First presented by Bellman and Åström (1970), structural analysis can help determine if the model structure affects parameter identifiability. Consider a nonlinear ordinary differential

equation of the form:

$$\frac{d}{dt}\mathbf{x}(t) = f(t, \mathbf{x}(t), \mathbf{q}), \qquad \mathbf{x}(0) = \mathbf{x}_0, \qquad y = g(t, \mathbf{x}(t), \mathbf{q}) \tag{4.1}$$

for t time, $\mathbf{x}(t)$ the model state varying according to time, $\mathbf{q}$ model parameters, $\mathbf{x}_0$ a vector of initial conditions, and y model predictions. The model and parameters are said to be identifiable if

$$g(t, \mathbf{x}, \mathbf{q}_1) = g(t, \mathbf{x}, \mathbf{q}_2) \iff \mathbf{q}_1 = \mathbf{q}_2. \tag{4.2}$$

If Equation (4.2) holds for the entire parameter space, then the parameters are said to be *globally* identifiable. Otherwise, if Equation (4.2) holds for $\mathbf{q}_2$ in a neighborhood of $\mathbf{q}_1$, then the parameters are said to be locally (or at-a-point) identifiable (Quaiser and Mönnigmann, 2009; Miao et al., 2011; McLean et al., 2012).

In the case of linear models, determining identifiability is a relatively straightforward process, using techniques such as Laplace transforms, power series expansion and similarity transformations (Quaiser and Mönnigmann, 2009; Miao et al., 2011). Two conditions for linear PBPK models are developed in Slob et al. (1997) to guarantee identifiability and estimability; however the authors are quick to point out that the methods do not extend nicely to nonlinear models. It is generally the case that assessing identifiability in nonlinear models is more difficult. Various techniques have been developed or extended to investigate the identifiability of a nonlinear model, including extensions of similarity transformations (Vajda et al., 1989), differential algebra (Ljung and Glad, 1994), and techniques based on the Implicit Function Theorem (Xia and Moog, 2003). Grewal and Glover (1976) showed that it is sufficient to show that if the linearized version of the ODE model was identifiable, then the nonlinear ODE model is also identifiable. However, the if the nonlinear ODE model is identifiable, it is not guaranteed that the linearized

version is also identifiable (McLean et al., 2012). The techniques listed above have been applied to the question of the implication of a non-identifiable linearized model (in both the local and global sense), and are summarized in (Miao et al., 2011; McLean et al., 2012). Although no conclusive method exists for large models, non-identifiability of the linearized model does suggest that the original model might be ill-formulated, at least in terms of the structure of model equations (McLean et al., 2012).

4.1.2 Estimability

Identifiability analysis provides a theoretical ground for practical identifiability (or estimability) analysis. If structural identifiability results suggest that a model is in fact unidentifiable, then it follows directly that the model is also inestimable (Miao et al., 2011). It is possible for portions of a model to be identifiable; if this is the case, then estimability analysis should only be run on the subset of these identifiable parameters.

Structural identifiability analysis relies heavily on the assumptions that the model structure is absolutely accurate and the measurements are exact (Miao et al., 2011). However, in practice, this is often not the case, and a parameter that is deemed identifiable *a-priori* may not be estimable due to data limitations. We say that a parameter is estimable if its value can be uniquely determined from available data. A good modeling technique is to combine structural and practical identifiability analysis to ensure reliable parameter estimations. If a parameter is determined to be structurally unidentifiable, then no amount of estimability analysis will change that. Conversely, if data is not amenable, an identifiable parameter may not be estimable under practical identifiability analysis. Miao et al. (2011) and McLean et al. (2012) detail several different methods for practical identifiability analysis; we will focus on techniques relating to sensitivity-based identifiability analysis.

4.1.3 Sensitivity-Based Identifiability Analysis

As methods for global identifiability and estimability of large models do not yet exist, we will focus our attention on local at-a-point analysis. A parameter may be difficult to estimate due to insensitivity, but it might also be correlated with other parameters, thus adding to the complexity of the estimation process. If there is a large correlation between a set of parameter estimates, then the elements of this set are likely to be indistinguishable. However, sensitivity analysis can help tease out the dependencies and provide a work-around to these complexities.

Consider again Equation (4.1):

$$\frac{d}{dt}\mathbf{x}(t) = f(t, \mathbf{x}(t), \mathbf{q}), \qquad y = g(t, \mathbf{x}(t), \mathbf{q}).$$

To determine the effects of parameters $\mathbf{q}$ on states $\mathbf{x}$, we can differentiate both sides of the model equations with respect to $\mathbf{q}$ to obtain a system of $n + mn$ differential equations:

$$\frac{d}{dt}\mathbf{x}(t) = f(t, \mathbf{x}(t), \mathbf{q}) \tag{4.3}$$

$$\frac{d}{dt}\frac{\partial \mathbf{x}}{\partial \mathbf{q}} = \frac{\partial f}{\partial \mathbf{x}}\frac{\partial \mathbf{x}}{\partial \mathbf{q}} + \frac{\partial f}{\partial \mathbf{q}}. \tag{4.4}$$

These sensitivity functions (Equation (4.4)) can be approximated using finite difference methods or computed exactly using Automatic Differentiation. Details about these methods are available in Appendix B.1. After computation of the sensitivity functions, we need to compare the sensitivity of a state to one parameter to the sensitivity of the state to that of another. We can accomplish this by first non-dimensionalizing the parameters

by multiplying by the parameter value,

$$\frac{\partial x_i}{\partial q_j}\, q_j. \tag{4.5}$$

Furthermore, we scale Equation (4.5) by dividing by the maximum value of the state variable x_i over $[t_0, t_f]$. Utilizing the maximum value rather than the value at a particular time t_k eliminates the possibility of dividing by zero, as all biological compartments are nonzero at t_0. To finalize the calculations, we can take the L^2 norm to achieve a *normalized ranking* of the sensitivity of each parameter over the time frame $[t_0, t_f]$:

$$C_{i,j} = \left\| \frac{\partial x_i}{\partial q_j}\, \frac{q_j}{\max x_i} \right\|_2^2 = \int_{t_0}^{t_f} \left| \frac{\partial x_i}{\partial q_j}\, \frac{q_j}{\max x_i} \right|^2 \, dt \in \mathbb{R}, \tag{4.6}$$

where t_0 and t_f are the initial and final time points, respectively. In practice, the integral can be approximated with the trapezoidal method. Parameters can be ranked in descending order of sensitivity to begin to yield ideas on identifiability on the system as a whole. In the simplest case, the larger the sensitivity ranking, the more influence a parameter will have on model behavior. This does not imply, however, that all sensitive parameters are uniquely estimable. For example, two parameters can have roughly the same normalized ranking, but cannot be identified from each other based on model data. We will see this is the case with kinetic parameters.

Typically speaking, a biological model will have many fewer output measurements than state variables. In the case of the vitamin D model, we have one output measurement, namely the concentration of 25(OH)D in the venous compartment. From this measurement, we are forced to estimate the other states in the model. To consider the sensitivities for the entire model through time with respect to the output measurements, we define a

sensitivity matrix $\mathbf{S}(\mathbf{q}, t)$. Recall $\mathbf{y} = [y_1(t), y_2(t), \ldots, y_p(t)]^T$ is a vector of outputs at time t, with parameters $\mathbf{q} = [q_1, q_2, \ldots, q_m]^T$ at time points $t = [t_1, t_2, \ldots t_n]$. The sensitivity of each system output with respect to the parameter vector is given by the $pn \times m$ matrix $\mathbf{S}(\mathbf{q}, t)$. This is a block matrix composed of p time dependent blocks $s_{i,\cdot}(\mathbf{q}, t)$ of size $n \times m$, as shown in Equation (4.7) (Beck and Arnold, 1977):

$$\mathbf{S}(\mathbf{q}, t) \equiv \begin{bmatrix} s_{1,\cdot}(\mathbf{q}, t) \\ s_{2,\cdot}(\mathbf{q}, t) \\ \vdots \\ s_{p,\cdot}(\mathbf{q}, t) \end{bmatrix}, \quad s_{i,\cdot}(\mathbf{q}, t) = \begin{bmatrix} s_{i,1}(\mathbf{q}, t_1) & \ldots & s_{i,m}(\mathbf{q}, t_1) \\ s_{i,1}(\mathbf{q}, t_2) & \ldots & s_{i,m}(\mathbf{q}, t_2) \\ \vdots & \ddots & \vdots \\ s_{i,1}(\mathbf{q}, t_n) & \ldots & s_{i,m}(\mathbf{q}, t_n) \end{bmatrix}, \tag{4.7}$$

where the entries of $s_{i,j}(\mathbf{q}, t)$ are called (normalized) sensitivity coefficients and indicate the influence of a perturbation in parameter q_j on the prediction of the output state y_i at time t_k. Formally, $s_{i,j}(\mathbf{q}, t_k)$ can be defined as follows:

$$s_{i,j}(t_k) = \frac{\partial y_i(t_k)}{\partial q_j} \; \frac{q_j}{\max y_i}. \tag{4.8}$$

A coefficient with a high value indicates that the output is highly sensitive to changes in a particular parameter; this is an indication (but not a guarantee) that the parameter is important to system behavior. Conversely, a parameter with a small sensitivity coefficient is a candidate for an unidentifiable parameter (Quaiser and Mönnigmann, 2009; Miao et al., 2011). Note that $\mathbf{S}(\mathbf{q}, t)$ requires evaluation at a particular point $\mathbf{q}$ in parameter space; thus, we consider this technique to be applicable only to at-a-point identifiability analysis. In addition to giving an idea of the strength of a perturbation of a parameter on model output, the columns of the sensitivity matrix can indicate correlation of parameters through linear dependencies.

Fisher Information Matrix

The sensitivity matrix $\mathbf{S}(\mathbf{q}, t)$ is a local result, evaluated at a fixed point in the parameter space $\mathbf{q}_0$. Let $\Delta\mathbf{q} = \mathbf{q} - \mathbf{q}_0$ be a small perturbation from the initial point $\mathbf{q}_0$. We expect a small change in the output function $\Delta\mathbf{y} = \mathbf{y}(t, \mathbf{q}) - \mathbf{y}(t, \mathbf{q}_0)$. Utilization of the chain rule gives the following relation:

$$\Delta\mathbf{y} = \mathbf{S}\Delta\mathbf{q}. \tag{4.9}$$

A model is said to be locally estimable if the modified *Fisher Information Matrix* (FIM) $\mathbf{S}^T\mathbf{S}$ has full rank, i.e. $rank(\mathbf{S}^T\mathbf{S}) = j$ where j is the number of model parameters. If the FIM does not have full rank, then not all parameters are estimable (McLean et al., 2012). Linearly dependent columns of the FIM imply that change in system output through variation of a parameter, say p_k, can be compensated by changing other parameter values $p_j, j \neq k$ as well. Note that the FIM can be written as $\mathbf{S}^T\Sigma\mathbf{S}$ where Σ is a square weighting matrix reflecting measurement noise information (McLean et al., 2012). For the purposes of the vitamin D PBPK model, we assume that Σ is the identity matrix since we have little information about measurement error.

Often in practice, calculation of the rank of $\mathbf{S}^T\mathbf{S}$ is fraught with numerical uncertainties. The condition number of the matrix, namely the ratio of the largest eigenvalue to the smallest eigenvalue, can be used to assess estimability; however, singular or nearly-singular matrices have large condition numbers, which can play into numerical computation difficulties. We will instead direct our attention to a separate method for assessing estimability of parameters based on the eigenvalues of the FIM.

Eigenvalue Method

Several methods have been proposed that use the sensitivity matrix $\mathbf{S}(\mathbf{q}, t)$; a detailed summary is presented in Miao et al. (2011). We will focus primarily on the eigenvalue method, first proposed by Vajda et al. (1989) and further discussed by Quaiser and Mönnigmann (2009).

The original eigenvalue method proposed by Vajda et al. (1989) focused on minimizing dependencies between eigenvectors arising due to special parameter combinations of the form p_1/p_2 or $p_1 p_2$ in the model. The first step entailed lumping these combinations into a new parameter, e.g. $p_3 = p_1 p_2$, and recalculating the eigenvalues and eigenvectors for the new system. These steps were repeated until the smallest eigenvalue was sufficiently large. The downfall of this technique is that it requires manual inspection of the eigenvalues and eigenvectors of the FIM and choices regarding lumping of the parameters, thus making this method infeasible for large systems.

To combat the limitations of this technique, Quaiser and Mönnigmann (2009) developed an automatic, algorithmically-based, method to discern interrelationships between parameters. This method amounts to calculating $\mathbf{S}^T\mathbf{S}$ and testing for positive-definiteness at a particular point $\mathbf{q}$ in parameter space. Positive-definiteness implies that all the eigenvalues are strictly positive; the authors in (Quaiser and Mönnigmann, 2009) utilize a lower bound ε on the eigenvalues of $\mathbf{S}^T\mathbf{S}$ to determine if the model is estimable. Eigenvalues that are smaller than ε indicate the presence of inestimable parameters. Parameters that are least estimable (elements in the eigenvectors corresponding to eigenvalues less than ε) are sequentially fixed to literature values and the matrix $\mathbf{S}^T\mathbf{S}$ is recalculated excluding the fixed parameter. The resulting set of parameters from utilization of this algorithm are parameters that are identifiable (since estimability analysis is only valid for identifiable

parameters) and estimable.

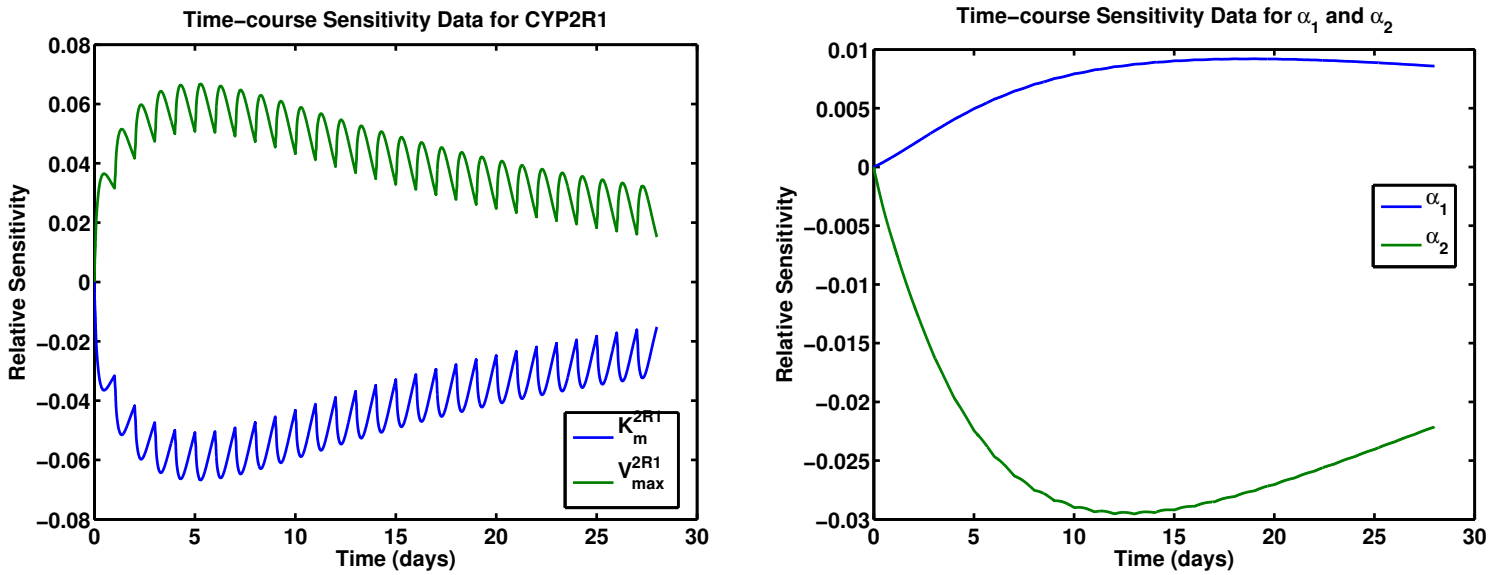
Visual Inspection

A final method of determining the sensitivity of a parameter is through visual inspection of the sensitivity functions. The normalized sensitivity of output state i with respect to parameter j over time is given by the function

$$S_{i,j}(t) = \frac{\partial y_i(t)}{\partial q_j} \frac{q_j}{\max y_i}. \tag{4.10}$$

These functions correspond to columns of the matrix $s_{i,:}(\mathbf{q}, t)$ in Equation (4.7). Note that nondimensionalization and scaling is essential when comparing sensitivities across parameters to ensure that apparently small sensitivity values are not a product of poor choice of units for parameters or response variables.

Sensitivity coefficients calculated using Equation (4.6) give an overall view of the sensitivity of a parameter; however, information about the sensitivity of a parameter over time is lost. This may lead to masking of interesting dynamics that may indicate particular time points or parameters of interest. Therefore, it may be important to assess time-course behavior of parameters that appear essential to the model but do not show up as sensitive or estimable, but have an effect on the model under empirical testing. Figure 4.1 shows the time-course sensitivities of four parameters: the kinetic pair for CYP2R1 and the pair (α_1, α_2) relating to the amount of p450 enzyme available. If the sensitivities for the kinetic enzymes were added together, the resulting curve would very nearly lie on the zero-axis. This indicates that the columns in $\mathbf{S}^T\mathbf{S}$ corresponding to the CYP2R1 kinetic parameters are linearly dependent. In contrast, the sensitivities for α_1 and α_2 do not show the same correlation, indicating the columns are not linearly



(a) K_m^{2R1} and V_{max}^{2R1}

(b) α_1 and α_2

Figure 4.1: Comparison of sensitivity of parameters over time. Kinetic parameters for CYP2R1 are shown on the left and α_1, α_2 on the right.

dependent. It is hasty to utilize only visual inspection to determine identifiability of parameters; two columns may not be linearly dependent, but with the addition of other columns (parameter sensitivities), a previous linearly independent column may become dependent given the basis for $\mathbf{S}^T\mathbf{S}$. For this reason, we choose to only use visual inspection as a graphical way of verifying the identifiability results.

4.2 Parameter Subset Selection

Before optimization of an arbitrary set of parameters, these parameters must be separated into those which are identifiable and sensitive, and those that are not. We use the following procedure:

1. Calculate the sensitivity matrix $\mathbf{S}$ and the total sensitivity of each parameter. Using the total sensitivity, determine the identifiable parameters.

2. Construct the modified Fisher Information Matrix $\mathbf{S}^T\mathbf{S}$ and utilize the eigenvalue method to decide which identifiable parameters are also estimable.

3. Apply optimization techniques to this subset of parameters.

4.2.1 Parameter Identifiabilty

Perhaps the easiest first step in identifiability and estimability analysis is to determine rankings of parameters based on sensitivity analysis. In the experimental design, three doses–0μg, 5μg, and 10μg–are given on a daily basis. Figure 4.2 shows the total sensitivity ranking of the parameters calculated utilizing Equation (4.6) and a 10μg dose. Note that the dose choice is arbitrary, as sensitivities for the other doses have the same overall

structure as the 10μg dose (results not shown). We chose to use the highest experimental dose for analysis to tease out the behavior of the smallest sensitivities.

Parameters are ranked in order from most sensitive to least sensitive, with sensitivity values given in Table 4.2. Obvious breaks in the ranking of the parameters are indications of transitions between identifiable to unidentifiable parameters; some discretion must be used, as the first two breaks appear within the first four parameters.

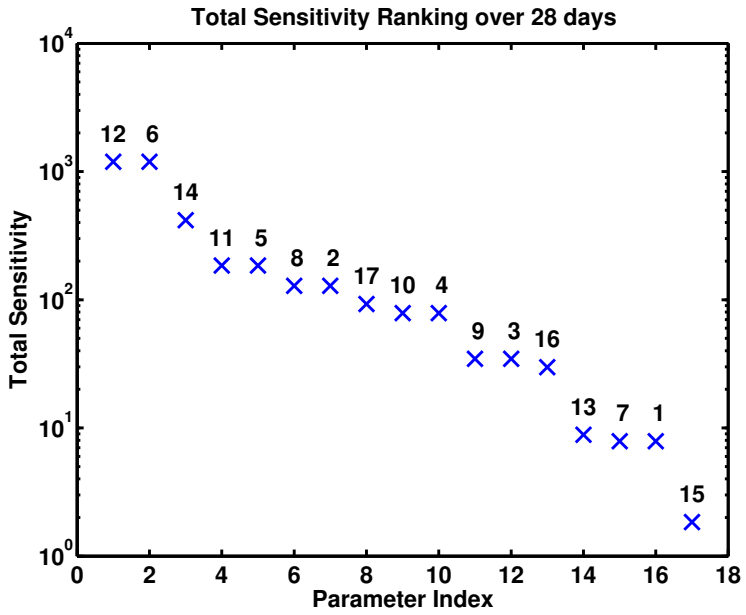


Figure 4.2: Total sensitivity ranking. Note the logarithmic scale for the sensitivity coefficients. Numbers above data points indicate the parameter.

The astute reader may notice that there are six instances in which pairs of parameters appear next to each other in Figure 4.2; in addition these pairs have nearly equal sensitivity coefficients, as given in Table 4.2. This is not a coincidence: the kinetic parameters K_m and V_{max} are correlated, as K_m is defined in terms of V_{max}. As discussed in Section 3.2.5, the slope of the metabolism function is roughly equal to $\frac{V_{max}}{K_m}$ for the dosing range of

Table 4.2: Sensitivity coefficients for **S**, ordered by relative sensitivity. The horizontal line separates identifiable from unidentifiable parameters.

Parameter #	Parameter	Sensitivity Index	Sensitivity Coefficient
12	V_{max}^{24A1b}	1	1191.4
6	K_m^{24A1b}	2	1191.3
14	P_{25}^{est}	3	417.5
11	V_{max}^{27A1b}	4	184.7
5	K_m^{27A1b}	5	184.7
8	V_{max}^{2R1}	6	128.5
2	K_m^{2R1}	7	128.5
17	α_2	8	92.4
10	V_{max}^{24A1}	9	78.7
4	K_m^{24A1}	10	78.7
9	V_{max}^{27B1}	11	34.6
3	K_m^{27B1}	12	34.6
16	α_1	13	29.8
13	P_D^{est}	14	8.9
7	V_{max}^{27A1}	15	7.9
1	K_m^{27A1}	16	7.9
15	$P_{1,25}^{est}$	17	1.8

the experiment. This indicates that although V_{max} and K_m for an enzyme may both be identifiable, they are most likely not uniquely estimable. For completeness of analysis, we will consider these parameters separately for the time being rather than lumping the parameters as suggested earlier in this chapter.

Note that in Figure 4.2 there are several fairly significant gaps in the ranking of

the parameters. These gaps occur after parameter indices two, three, ten, thirteen and sixteen. As gaps in the ranking yield clues as to where the division between identifiable and unidentifiable parameters lies, we must make an informed choice about which break to choose as the dividing line. The last two gaps occur between parameters with very low relative sensitivities; in a similar fashion, the first two gaps occur between parameters with relatively high sensitivity. This leaves the break between parameter indices ten and eleven as an optimal choice for identifiability separation.

The identifiability results from the spectrum of the information matrix are not surprising: parameters that are most structurally identifiable are directly related to 25(OH)D, which is the metabolite for which we have data. The least identifiable set of kinetic parameters is for CYP27A1 acting on vitamin D. As discussed previously, this may actually be the secondary pathway for CYP27A1, as it seems to prefer hydroxylated forms of the compound.

4.2.2 Parameter Estimability

As discussed previously, we are using the eigenvalue method to determine parameters that are both identifiable and estimable. A brief description of the algorithm is summarized from (Quaiser and Mönnigmann, 2009).

Let $p = \{p_1, p_2, \ldots, p_n\}$ denote the set of parameter values that are deemed identifiable from previous analysis and p^{lit} denote the associated literature values of these parameters. Let $\mathcal{S}$ denote the set of estimable parameters and $\mathcal{U}$ denote the set of inestimable parameters.

1. Initialize the model by choosing ε, setting $p^* = p^{lit}$ as an evaluation point. Set $\mathcal{S} = p$, and $\mathcal{U} = \emptyset$. For this analysis, we chose $\varepsilon = 10^{-8}$.

2. If $\mathcal{S}$ is empty, then stop. This implies that the model is not estimable with respect to any parameter.

3. Fix any parameters $p_k \in \mathcal{U}$ to their literature values. These parameters are deemed inestimable, and thus we can gain no further information by varying them.

4. Calculate the Fisher Information Matrix with respect to parameters $p_j \in \mathcal{S}$ and evaluate it at $p = p^*$. Find the eigenvalues λ_j and associated normalized eigenvectors v_j of the FIM. Assume these pairs are ordered such that $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_j$.

5. If $\lambda_1 \geq \varepsilon$, then stop. The model is estimable with respect to parameters $p_j \in \mathcal{S}$.

6. Otherwise, chose the largest element $|v_1^i|$ in the eigenvector v_1. This corresponds to the influence of the parameter p_i on the model. Remove p_i from the set $\mathcal{S}$ and add it to the set $\mathcal{U}$. This indicates that p_i is not an estimable parameter with respect to the model. Return to step 2.

The order in which parameters are removed from $\mathcal{S}$ determines the ranking of parameters from least to most estimable. After determining which parameters are identifiable from the model structure, we can apply the eigenvalue method to determine the most estimable parameters.

As expected, estimability analysis has indicated that only one of each identifiable kinetic parameter pair is estimable. The identifiable parameters that are deemed most estimable are indicated in Figure 4.3 by red circles.

4.2.3 Optimization Methods

Parameters that are deemed both identifiable and estimable are expected to have an influence on the model output. We may then optimize for these parameters, holding all

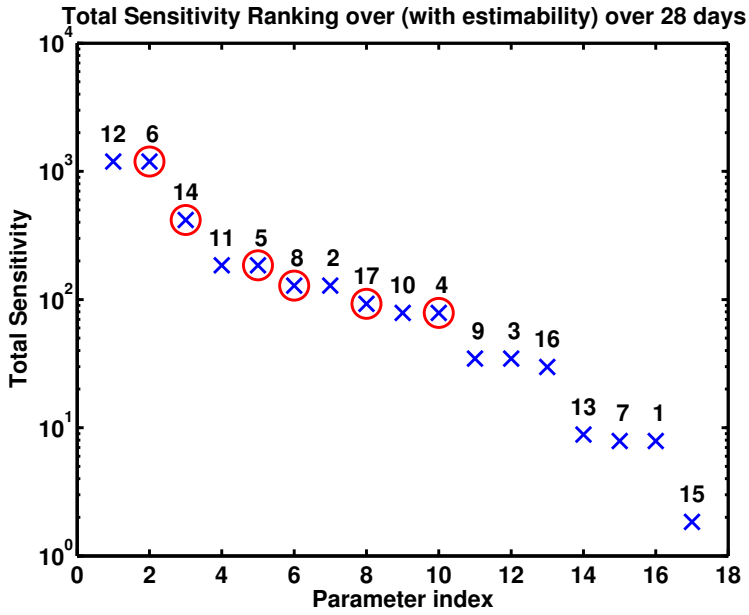


Figure 4.3: Ranked parameters utilizing the spectrum of the information matrix. Parameters that are both estimable and identifiable are circled in red.

others at fixed literature values, to determine a parameter combination that yields a "best" fit based on some cost function. In particular, the cost function that we use for this model is given by

$$\phi(\mathbf{q}) = \sum_{i=1}^{\gamma} \sum_{j=1}^{n} \left[f_i(t_j, \mathbf{x}(t_j), \mathbf{q}) - y_i(t_j) \right]^2, \tag{4.11}$$

with respect to $\mathbf{q}$, where $f_i(t_j, \mathbf{x}(t_j), \mathbf{q})$ is the model output and $y_i(t_j)$ are the measured data. In the case of the vitamin D model, there is only one set of measurement data, so $\gamma = 1$. In particular, the output state $y(t)$ corresponds to the state $f_{20}(t, \mathbf{x}(t), \mathbf{q})$ Thus, Equation (4.11) simplifies to

$$\phi(\mathbf{q}) = \sum_{j=1}^{n} \left[f_{20}(t_j, \mathbf{x}(t_j), \mathbf{q}) - y(t_j) \right]^2, \tag{4.12}$$

Various algorithms are available to minimize functions of this form; we chose to use the MATLAB function `fminsearch`, which utilizes the Nelder-Mead search algorithm, to find an initial local minimum. After an initial local minimum is found with a relatively good model fit to data, we switch algorithms to a gradient-based method (`lsqnonlin`) to speed up the convergence. Details about both types of methods can be found in Appendix B.

4.3 Parameter Confidence Intervals

The above techniques describe methods for choosing a parameter subset for estimation. To determine the effectiveness of different parameter estimation techniques, we can consider the standard errors of the resulting parameter estimations. To this end, we follow notation in (Banks and Tran, 2009) for the development of this idea.

Consider the case of n scalar longitudinal observations for a generic statistical model given by

$$Y_k \equiv f(t_k, \mathbf{q}_0) + \mathcal{E}_k, \quad k = 1, \dots n, \tag{4.13}$$

where $f(t_k, \mathbf{q}_0)$ is the model for the observations in terms of the state variables, and $\mathbf{q}_0 \in \mathbb{R}^p$ is the set of "true" parameter values. The observational errors, $\mathcal{E}_k$, are assumed to be independently and identically distributed (i.i.d.) random variables with mean zero and an unknown variance σ_0^2. This implies that the observations Y_k are i.i.d. random variables with mean $f(t_k, \mathbf{q}_0)$ and variance σ_0^2.

In the ordinary least squares (OLS) approach, the goal is to use realizations $\{y_k\}$ of the process $\{Y_k\}$ along with the model to determine a vector $\mathbf{q}^*$ where

$$\mathbf{q}^* = \arg\min J(q) = \sum_{k=1}^{n} \big(y_k - f(t_k, q)\big)^2. \tag{4.14}$$

Y_k is a random variable; this implies that the corresponding estimator $\mathbf{q}^*$ is also a random variable with a distribution called the sampling distribution. This sampling distribution characterizes all possible values that $\mathbf{q}^*$ could take on for a fixed data set of length n. The standard errors approximate how much variability is present across all realizations and provide a measure of the extent of uncertainty. This variability is dependent on both the estimator $\mathbf{q}^*$ and the sample size n.

The sample distribution can be estimated with a p-multivariate Gaussian distribution (with asymptotic convergence to distribution) with mean $\mathbf{q}_0$ and covariance matrix $\Sigma_0 = \sigma_0^2(S^T(\mathbf{q}_0)S(\mathbf{q}_0))^{-1}$. Here, $S(\mathbf{q}_0)$ is the local, regular sensitivity matrix evaluated at $\mathbf{q}_0$. For n large, the sampling distribution approximately satisfies

$$\mathbf{q}^*(Y) \sim \mathcal{N}_p(\mathbf{q}_0, \Sigma_0) \approx \mathcal{N}_p(\mathbf{q}_0, \sigma_0^2[S^T(\mathbf{q}_0)S(\mathbf{q}_0)]^{-1}). \tag{4.15}$$

Methods for computing the sensitivity matrix have been discussed previously in this chapter; we can approximate σ_0^2 by

$$\sigma_0^2 \approx \frac{1}{n-p}\sum_{k=1}^{n}\big(y_k - f(t_k, \mathbf{q}^*)\big)^2. \tag{4.16}$$

Standard errors are given by the square root of the diagonal of the covariance matrix Σ:

$$E_k(\mathbf{q}) = \sqrt{\Sigma_{kk}(\mathbf{q})} \text{ for } k = 1, \ldots n. \tag{4.17}$$

The confidence intervals at the $100(1-\alpha)\%$ level are defined in terms of these errors:

$$P\{\mathbf{q} - t_{1-\alpha/2}E_k(\mathbf{q}) < \mathbf{q}_0 < \mathbf{q} + t_{1-\alpha/2}E_k(\mathbf{q})\} = 1 - \alpha, \tag{4.18}$$

for $\alpha \in [0, 1]$ and $t_{1-\alpha/2} \in \mathbb{R}^+$. For small values of α, e.g. 0.05, the values for $t_{1-\alpha/2}$ are computed from the Student's t distribution with $n - p$ degrees of freedom.

We will present confidence intervals on our optimized parameters in Chapter 5.

4.4 Parameter Estimation

We utilized the following algorithm to determine the optimized parameter values for the vitamin D model.

1. Calculate the sensitivity coefficients and the sensitivity ranking of parameters at a nominal starting point.

2. Determine parameters that are identifiable using the total sensitivity coefficients of $\mathbf{S}^T\mathbf{S}$; from these, utilize the eigenvalue method to determine which of these are estimable.

3. Partition $\mathbf{q} = \{\mathbf{q}_I, \mathbf{q}_U\}$ such that parameters in $\mathbf{q}_I$ are both identifiable and sensitive utilizing methods discussed previously. Parameters in $\mathbf{q}_U$ are unidentifiable by the model.

4. Determine if any parameters in $\mathbf{q}_U$ warrant special consideration by investigating time-course sensitivities. Remove these parameters from $\mathbf{q}_U$ and add them to the set $\mathbf{q}_I$.

5. Set all parameters in $\mathbf{q}_U$ to their associated literature values. Parameters in $\mathbf{q}_I$ will be optimized for, with kinetic search regions constrained by available literature data (found in Table A.4).

6. Determine a local minimum for parameters in $\mathbf{q}_I$ by minimizing Equation (4.12) utilizing a Nelder-Mead optimizer, `fminsearch`.

7. Utilize a gradient-based algorithm, `lsqnonlin`, to speed up the convergence.

8. Display model output.

Chapter 5 discusses the results from application of this parameter estimation algorithm.

Chapter 5: Model Results

Modeling, especially modeling of biological systems, is an iterative process. In the case of biological modeling, an initial attempt may fit portions of the data but grossly overestimate or underestimate other portions of the data. It is at this point that more biology must be investigated and evaluated for inclusion into the model. We see this is the case with the vitamin D PBPK model. Each section of this chapter discusses the progression through the different levels of complexity of the PBPK model and how each additional idea was incorporated and the resulting model fit.

5.1 Initial Model Structure

In the initial model, we strived to keep the mathematics as simple as possible while incorporating the majority of known biology. To this end, we needed to include not only the major pathways of metabolism, but also the secondary pathways as well (namely, CYPs 27A1 and 24A1 acting on the compound 25(OH)D). The resulting fit to the data is

shown in Figure 5.1. The spikes in each model come from the daily dosing of vitamin D through dietary consumption. Clearly, the zero-dosing scheme has a relatively good fit, while the model vastly over-predicts the other two dosing schemes.

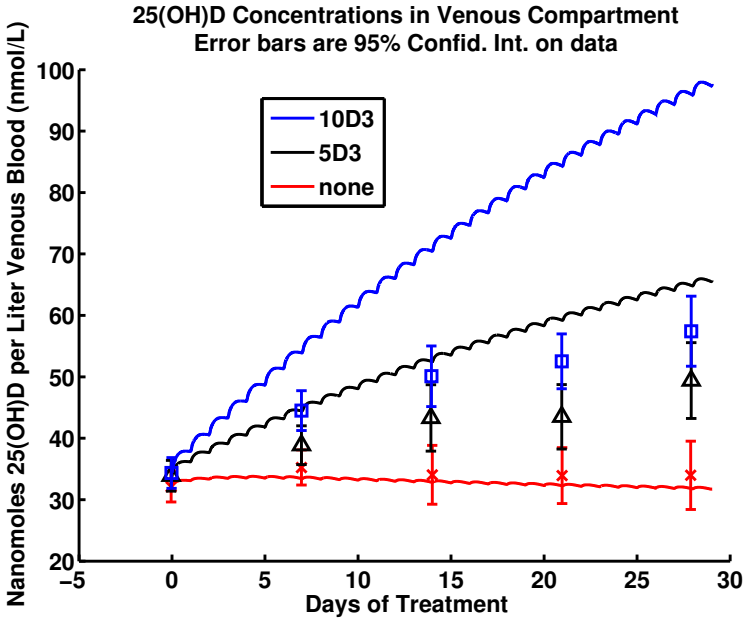


Figure 5.1: Initial model results. This includes only the primary and secondary pathways.

A positive aspect of this initial model setup is that the model fits the zero-dose data very well. This indicates that the majority of our parameters relating to steady-state values are approximately accurate as the initial state conditions were also derived from a zero-dose scheme (as discussed in Appendix A.2). This is an aspect of the model that we would like to preserve, so changes to the model must occur in relationship to increasing the dose. Note that the zero-dose prediction does decrease slightly over time; this may be due to the individual variations in typical vitamin D consumption by study participants.

5.2 Dynamic Adipose Partition Coefficients

The next iteration of the model explored the effects of a daily dosing scheme on the adipose partition coefficients. Recall from Section 3.2.4 that a partition coefficient is calculated with the assumption of steady-state conditions. However, with the daily dosing of vitamin D, we may not have a steady-state environment. Equation (3.8) defines the adipose tissue partition coefficient in terms of the concentration of the compound in the venous compartment compared to the estimated steady state concentration. Figure 5.2 illustrates how the value of the adipose tissue partition coefficient changes with respect to time and dosing amount.

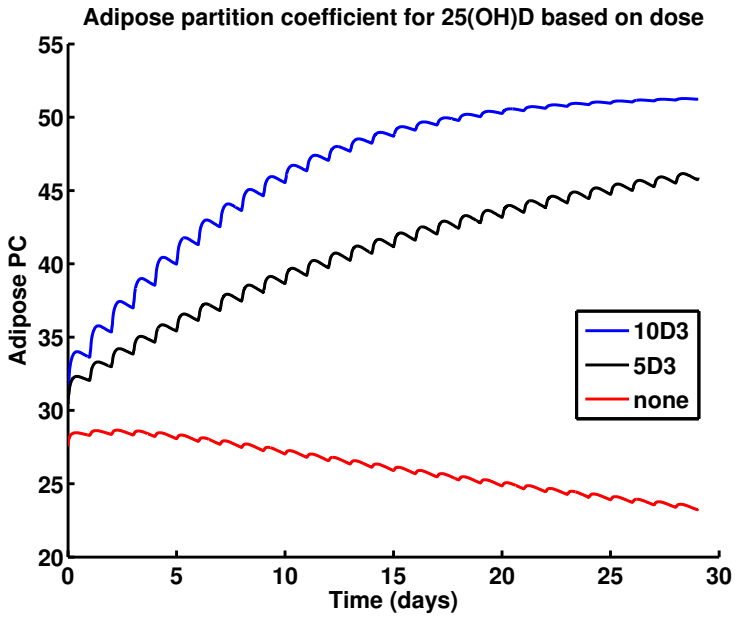


Figure 5.2: Time-course display of the values of the dynamic adipose partition coefficients.

All doses start with roughly the same adipose PC value; however, we see a strong

increase in the value of the partition coefficient as time progresses with the addition of a dose of vitamin D to the diet. This corresponds to sequestering of the metabolite 25(OH)D in the adipose tissue. Conversely, the zero-dosing scheme yields a decreasing value of the adipose tissue PC; this indicates donating behavior of the adipose tissue in response to falling levels of 25(OH)D throughout the rest of the system. The adipose partition coefficients for the other compounds display a similar behavior.

In terms of the model output, we expect that varying the adipose PC will have little effect on the model for the zero-dosing scheme. This is because we expect little change from steady-state levels of 25(OH)D since there is no appreciable change in the dietary intake of vitamin D by the subjects. Conversely, we expect that dynamic adipose partition coefficients will show a demonstrable effect on the model output for the other two dosing schemes, most apparent in the 10μg dosing scheme. Indeed, this is the case, as shown in Figure 5.3.

With the addition of the dynamic adipose partition coefficients, we see a strong improvement in the fit for all doses over the course of the first 14 days, and continued improvement for the 5μg dosing over the remainder of the experiment. Very little has changed for the zero-dosing scheme, as anticipated, and the 5μg dosing predictions fit very tightly to the available data. The model predictions for the 10μg dosing indicate a trend with a concave-upward shape; this suggests that over a longer time period, the 10μg dosing model may demonstrate the same ill-fitting issues as seen in the 5μg dosing without the addition of the dynamic adipose PCs. Note also that the maximum concentration of the 5μg dosing is roughly equal to the concentration of the 10μg dosing on day 7.

To address this issue, the next adaptation of the model would ideally affect the model predictions only at high concentrations of 25(OH)D in the venous compartment, regardless of dosing scheme.

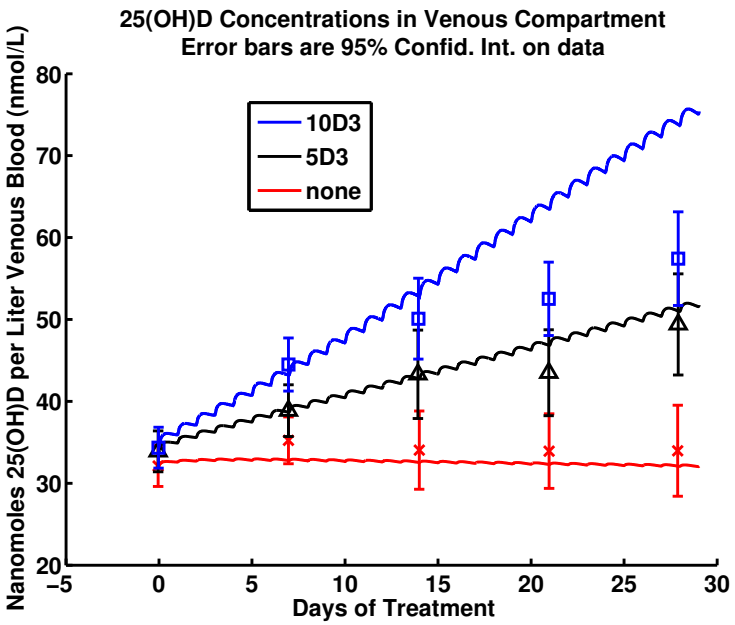


Figure 5.3: Model results with the addition of dynamic adipose partition coefficients.

5.3 Final Model and Optimized Parameters

The second adaption of the vitamin D PBPK model incorporated transcriptional regulation of two enzymes involved in the cascade: CYPs 27B1 and 24A1. Recall, as discussed in Section 2.2.5, the concentration of 1,25(OH)2D regulates the physical availability of these two enzymes by controlling the amount of enzyme produced. We incorporate this phenomenon into the PBPK model by allowing the amount of available enzyme to change according to methods discussed previously.

Adjusting the amount of enzyme available serves to effectively decrease the concentration of 25(OH)D in the venous compartments at high concentration levels while sacrificing little accuracy at lower concentrations. Figure 5.4 displays the final results.

Throughout the modeling process, we assumed several classes of parameters were

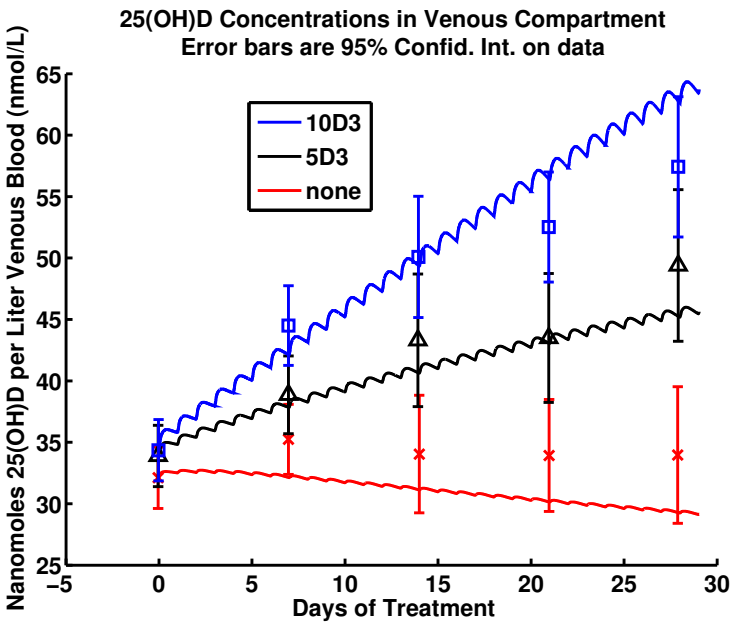


Figure 5.4: Model results with the addition of dynamic adipose partition coefficients as well as dynamic enzyme amounts.

constant and obtained through literature search. These parameter classes included compartment volumes, blood flow rates, and partition coefficients (with the exception of adipose tissue). These values can be found in Table 3.2. The remaining parameters fall into two classes: kinetic parameters (including the amount of enzyme available for CYPs 27B1 and 24A1) and the adipose partition coefficients. Table 5.1 displays these values. Note that an asterisk indicates that the parameter has been optimized for, and contains the 95% confidence interval. Non-optimized parameters are set to literature values, and do not have associated confidence intervals.

With the advancement of understanding of the metabolic cascade for vitamin D, we expect to be able to incorporate more detail in the model to truly capture the short and long term aspects of the data, as well as increase our confidence in literature and optimized values.

Table 5.1: Final optimized parameter values. Bold-face entries with an asterisk indicate optimized parameters, with *Lower* and *Upper* denoting the 95% confidence intervals (CI) on an optimized parameter.

Parameter	K_m			V_{max}		
	Lower CI	*Value*	*Upper CI*	*Lower CI*	*Value*	*Upper CI*
CYP27A1	–	3.20	–	–	0.27	–
CYP2R1	–	0.45	–	0.988	**0.99***	1.002
CYP27B1	–	2.70	–	–	3.90	–
CYP24A1	0.116	**0.12***	0.123	–	0.105	–
CYP27A1b	3.69	**3.89***	4.10	–	0.021	–
CYP24A1b	0.368	**0.37***	0.372	–	0.12	–

Parameter	*Lower CI*	*Value*	*Upper CI*
P_D^{est}	–	31.33	–
P_{25}^{est}	40.54	**40.65***	40.77
$P_{1,25}^{est}$	–	31.9	–
α_1	–	10.0	–
α_2	13.32	**14.0***	14.71

Chapter 6: Conclusion and Future Directions

In this work, we have constructed a human PBPK model that describes the distribution of the metabolic pathway of vitamin D simultaneously over three oral dosing levels. In contrast to typical PBPK models, the dosing compound, vitamin D, is not a naïve substance to the human body. In fact, it is an internally-created compound, produced through sunlight exposure and stored in various forms for utilization in darker months. Therefore, the traditional assumption of zero or negligible amounts of the dose initially present in each compartment is violated; we developed a novel technique to estimate the amount of vitamin D and its metabolites in each compartment prior to the start of the experiment. This allows the system to avoid the equilibrating process that would otherwise skew the model predictions.

The vitamin D PBPK model is not a simple model. With thirty differential equations constructing the backbone of the model, six kinetic equations, and thirty equations predicting the partition coefficients, an exact solution is not possible. We are still able to show existence and uniqueness of the model and its modifications by way of Lipschitz

continuity. As parameter estimation relies heavily on local sensitivity and estimability analysis, we utilize various methods to select pertinent parameters for optimization. These methods include construction and analysis of the Fisher Information Matrix to determine identifiability and use of an eigenvalue-based method for parameter subset selection. After selection, we use several optimization methods to determine values that lead to optimal model fits while satisfying biological constraints on parameter values.

Through incorporating a variety of biological phenomena, this model makes several new contributions to PBPK modeling. These include: a novel method of calculating initial conditions in a system with previous exposure to the substrates; concentration-dependent partition coefficients reflecting a lack of steady-state conditions in the adipose tissue compartment; and coupling of metabolic compounds through six kinetic pathways with varying amounts of enzymes available through concentration-dependent transcriptional regulation. Although these contributions push the boundary of PBPK modeling, we feel confident that they are justified in describing the phenomena we observe.

As public and clinical interest in vitamin D continues to grow, we expect a variety of new experiments will help shape our understanding of the underlying metabolic behavior of the vitamin D pathway. With an enhanced understanding of the relevant behavior of the four enzymes we have included—or the revelation of new enzymes or pathways involved—we can refine the equations and parameters describing conversion of one compound to another. This may be essential with the exploration of alternate dosing schemes, such as high level dosing or alternate routes. New techniques are being developed to quantitatively predict sunlight exposure; these results can be incorporated into this PBPK model for datasets utilizing exposure as a dosing route. As the interplay between internally-produced and externally-administered vitamin D is investigated, we hope to see several exciting predictions come arise from our PBPK model.

REFERENCES

Adams, J.S., Hewison, M.: Update in vitamin D. Journal of Clinical Endocrinology and Metabolism **95**, 471–478 (2010). DOI 10.1210/jc.2009-1773

Aiba, I., Yamasaki, T., Shinki, T., Izumi, S., Yamamoto, K., Yamada, S., Terato, H., Ide, H., Ohyama, Y.: Characterization of rat and human CYP2J enzymes as vitamin D 25-hydroxylases. Steroids **71**(10), 849–856 (2006). DOI 10.1016/j.steroids.2006.04.009

Avioli, L., Lee, S., McDonald, J., Lund, J., DeLuca, H.: Metabolism of vitamin D3-3H in human subjects: distribution in blood, bile, feces, and urine. Journal of Clinical Investigation **46**(6), 983–992 (1967). DOI 10.1172/JCI105605

Banks, H.T., Tran, H.T.: Mathematical and experimental modeling of physical and biological processes. Chapman and Hall/CRC (2009)

Battault, S., Whiting, S., Peltier, S., Sadrin, S., Gerber, G., Maixent, J.: Vitamin D metabolism, functions and needs: from science to health claims. European Journal of Nutrition **52**(2), 429–441 (2013). DOI 10.1007/s00394-012-0430-5

Beck, J.V., Arnold, K.J.: Parameter estimation in engineering and science, vol. 8. Wiley New York (1977)

Bellman, R., Åström, K.: On structural identifiability. Mathematical Biosciences **7**(3), 329–339 (1970). DOI 10.1016/0025-5564(70)90132-X

Bikle, D.: Nonclassic actions of vitamin D. Journal of Clinical Endocrinology and Metabolism **94**(1), 26–34 (2009). DOI 10.1210/jc.2008-1454

Bikle, D., Gee, E., Halloran, B., Haddad, J.: Free 1, 25-dihydroxyvitamin D levels in serum from normal subjects, pregnant subjects, and subjects with liver disease. Journal of Clinical Investigation **74**(6), 1966–1971 (1984). DOI 10.1172/JCI111617

Bills, C.E.: Antiricketic substances: VI. the distribution of vitamin D, with some notes on its possible origin. Journal of Biological Chemistry **72**(2), 751–758 (1927). URL `http://www.jbc.org/content/72/2/751.short`

Bouillon, R.: The vitamin D binding protein DBP, vol. 3, chap. 5, pp. 57–72. Academic Press (2011)

Carmichael, G.R., Sandu, A., Potra, F.A.: Sensitivity analysis for atmospheric chemistry models via automatic differentiation. Atmospheric Environment **31**(3), 475–489 (1997). DOI 10.1016/S1352-2310(96)00168-9

Chandler, P.T., Cragle, R.G.: Investigation of calcium, phosphorus and vitamin D3 relationships in rats by multiple regression techniques. The Journal of Nutrition **78**(1), 28–36 (1962). URL `http://jn.nutrition.org/content/78/1/28.short`

Cheng, J., Levine, M., Bell, N., Mangelsdorf, D., Russell, D.: Genetic evidence that the human CYP2R1 enzyme is a key vitamin D 25-hydroxylase. Proceedings of the National Academy of Sciences of the United States of America **101**(20), 7711–7715 (2004). DOI 10.1073/pnas.0402490101

Cheng, J., Motola, D., Mangelsdorf, D., Russell, D.: De-orphanization of cytochrome P450 2R1. Journal of Biological Chemistry **278**(39), 38,084–38,093 (2003). DOI 10.1074/jbc.M307028200

Chun, R., Peercy, B., Adams, J., Hewison, M.: Vitamin D binding protein and monocyte response to 25-hydroxyvitamin D and 1,25-dihydroxyvitamin D: Analysis by mathematical modeling. PLoS one **7**(1), e30,773 (2012). DOI 10.1371/journal.pone.0030773

Clement, B., Mau, S., Deters, S., Havemeyer, A.: Hepatic, extrahepatic, microsomal, and mitochondrial activation of the N-hydroxylated prodrugs benzamidoxime, guanoxabenz, and Ro 48-3656 ([[1-[(2s)-2-[[4-[(hydroxyamino) iminomethyl] benzoyl] amino]-1-oxopropyl]-4-piperidinyl] oxy]-acetic acid). Drug Metabolism and Disposition **33**(11), 1740–1747 (2005). DOI 10.1124/dmd.105.005249

Clewell, R.A., Clewell III, H.J.: Development and specification of physiologically based pharmacokinetic models for use in risk assessment. Regulatory Toxicology and Pharmacology **50**(1), 129–143 (2008). DOI 10.1016/j.yrtph.2007.10.012

Coburn, J.W., Koppel, M.H., Brickman, A.S., Massry, S.G.: Study of intestinal absorption of calcium in patients with renal failure. Kidney International **3**(4), 264–272 (1973). DOI 10.1038/ki.1973.40

Cuello, W., Janes, T., Jessee, J., Venecek, M., Sawyer, M., Eklund, C., Evans, M.: Physiologically based pharmacokinetic (PBPK) modeling of metabolic pathways of bromochloromethane in rats. Journal of Toxicology (2012). DOI 10.1155/2012/629781

De Buck, S., Sinha, V., Fenu, L., Nijsen, M., Mackie, C., Gilissen, R.: Prediction of human pharmacokinetics using physiologically based modeling: a retrospective analysis of 26 clinically tested drugs. Drug Metabolism and Disposition **35**(10), 1766–1780 (2007). DOI 10.1124/dmd.107.015644

Dennis, J., Schnabel, R.: Numerical methods for unconstrained optimization and nonlinear equations, vol. 16. Philadelphia: SIAM (1996)

D'Errico, J.: fminsearchbnd, fminsearchcon. MATLAB Central File Exhange (2006). URL http://www.mathworks.com/matlabcentral/fileexchange/8277-fminsearchbnd-fminsearchcon. Retrieved May 20, 2013

Deurenberg, P., Weststrate, J.A., Seidell, J.C., et al.: Body mass index as a measure of body fatness: age-and sex-specific prediction formulas. British Journal of Nutrition **65**(2), 105–114 (1991). DOI 10.1079/BJN19910073

Diffey, B.L.: Modelling the seasonal variation of vitamin D due to sun exposure. British Journal of Dermatology **162**, 1342–1348 (2010). DOI 10.1111/j.1365-2133.2010.09697.x

Dunn, J.: Computer simulation of vitamin D transport. Annals of the New York Academy of Sciences **538**(1), 69–76 (1988). DOI 10.1111/j.1749-6632.1988.tb48851.x

Dusso, A.S., Brown, A.J., Slatopolsky, E.A.: Chapter 76 - vitamin d and renal failure. In: Feldman, D., Pike, J.W., Glorieux, F.H. (eds.) Vitamin D (2), 2 edn., pp. 1313 – 1338. Academic Press, Burlington (2005). DOI 10.1016/B978-012252687-9/50079-6

Feldman, H.: Mathematical theory of complex ligand-binding systems of equilibrium: some methods for parameter fitting. Analytical Biochemistry **48**(2), 317–338 (1972). DOI 10.1016/0003-2697(72)90084-X

Fisk, C., Theobald, H., Sanders, T.: Fortified malted milk drinks containing low-dose ergocalciferol and cholecalciferol do not differ in their capacity to raise serum 25-hydroxyvitamin D concentrations in healthy men and women not exposed to UV-B. Journal of Nutrition **142**(7), 1286–1290 (2012). DOI 10.3945/jn.111.156166

Fraser, D.R.: The physiological economy of vitamin D. The Lancet **321**(8331), 969 – 972 (1983). DOI 10.1016/S0140-6736(83)92090-1. ¡ce:title¿Originally published as Volume 1, Issue 8331¡/ce:title¿

Fukushima, M., Suzuki, Y., Tohira, Y., Matsunaga, I., Ochi, K., Nagano, H., Nishii, Y., Suda, T.: Metabolism of 1α-hydroxyvitamin D3 to 1α, 25-dihydroxyvitamin D3 in perfused rat liver. Biochemical and Biophysical Research Communications **66**(2), 632–638 (1975). DOI 10.1016/0006-291X(75)90556-2

Ghazarian, J.G., Jefcoate, C.R., Knutson, J.C., Orme-Johnson, W.H., DeLuca, H.F.: Mitochondrial cytochrome p450 a component of chick kidney 25-hydroxycholecalciferol-1α-hydroxylase. Journal of Biological Chemistry **249**(10), 3026–3033 (1974). DOI 10.1016/0006-291X(75)90556-2

Grewal, M., Glover, K.: Identifiability of linear and nonlinear dynamical systems. Automatic Control, IEEE Transactions on **21**(6), 833–837 (1976). DOI 10.1109/TAC.1976.1101375

Griewank, A.: On automatic differentiation. Mathematical Programming: recent developments and applications **6**, 83–107 (1989). URL http://info.mcs.anl.gov/pub/tech_reports/reports/P10.pdf

Guo, Y.D., Strugnell, S., Back, D.W., Jones, G.: Transfected human liver cytochrome P-450 hydroxylates vitamin D analogs at different side-chain positions. Proceedings of the National Academy of Sciences **90**(18), 8668–8672 (1993). URL http://www.pnas.org/content/90/18/8668.full.pdf+html

Gupta, R.P., Hollis, B.W., Patel, S.B., Patrick, K.S., Bell, N.H.: CYP3A4 is a human microsomal vitamin D 25-hydroxylase. Journal of Bone and Mineral Research **19**(4), 680–688 (2004). DOI 10.1359/JBMR.0301257

Havard, R.E., Hoyle, J.C.: Vitamin D in adults: Its effect on the calcium and inorganic phosphate of the blood. Biochemical Journal **22**(3), 713–716 (1928). URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1252175/pdf/biochemj01138-0107.pdf

Heaney, R.P., Davies, K.M., Chen, T.C., Holick, M.F., Barger-Lux, M.J.: Human serum 25-hydroxycholecalciferol response to extended oral dosing with cholecalciferol. American Journal of Clinical Nutrition **77**, 204–210 (2003). URL http://ajcn.nutrition.org/content/77/1/204.full.pdf+html

Henry, H.L.: Regulation of vitamin D metabolism. Best Practice and Research Clinical Endocrinology and Metabolism **25**, 531–541 (2011). DOI 10.1016/j.beem.2011.05.003

Hewison, M.: Vitamin D and the immune system: New perspectives on an old theme. Endocrinology and metabolism clinics of North America **39**(2), 365–379 (2010). DOI 10.1016/j.ecl.2010.02.010

Holick, M.F.: Photosynthesis of vitamin D in the skin: effect of environmental and life-style variables. Federation Proceedings **46**, 1876–1882 (1987)

Holick, M.F.: Vitamin D: A millenium perspective. Journal of Cellular Biochemistry **88**, 296–307 (2003). DOI 10.1002/jcb.10338

Holick, M.F.: Vitamin D deficiency. New England Journal of Medicine **357**(3), 266–281 (2007). DOI 10.1056/NEJMra070553

Holick, M.F., Biancuzzo, R.M., Chen, T.C., Klein, E.K., Young, A., Bibuld, D., Reitz, R., Salameh, W., Ameri, A., Tannenbaum, A.D.: Vitamin D2 is as effective as vitamin D3 in maintaining circulating concentrations of 25-hydroxyvitamin D. Journal of Clinical Endocrinology & Metabolism **93**(3), 677–681 (2008). DOI 10.1210/jc.2007-2308

Hollander, D., Muralidhara, K., Zimmerman, A.: Vitamin D-3 intestinal absorption in vivo: Influence of fatty acids, bile salts, and perfusate pH on absorption. Gut **19**(4), 267–272 (1978). DOI 10.1136/gut.19.4.267

Houghton, L.A., Vieth, R.: The case against ergocalciferol (vitamin D2) as a vitamin supplement. American Journal of Clinical Nutrition **84**, 694–697 (2006). URL `http://ajcn.nutrition.org/content/84/4/694.full.pdf+html`

Hrycay, E.G., Bandiera, S.M.: Cytochrome P450 Enzymes, pp. 627–696. John Wiley & Sons, Inc. (2007)

Hume, E.M., Lucas, N.S., Smith, H.H.: On the absorption of vitamin D from the skin. Biochemical Journal **21**(2), 362–367 (1927). URL `http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1251921/pdf/biochemj01145-0111.pdf`

Jones, G.: Pharmacokinetics of vitamin D toxicity. American Journal of Clinical Nutrition **88**(2), 582S–586S (2008). URL `http://ajcn.nutrition.org/content/88/2/582S.full.pdf+html`

Jones, G.: Metabolism and catabolism of vitamin D, its metabolites and clinically relevant analogs. In: Holick, M.F. (ed.) Vitamin D, Nutrition and Health, 2 edn., pp. 99–134. Humana Press (2010)

Jones, G., Prosser, D.E., Kaufmann, M.: 25-hydroxyvitamin D-24-hydroxylase (CYP24A1): Its important role in the degradation of vitamin D. Archives of Biochemistry and Biophysics **523**(1), 9–18 (2012). DOI 10.1016/j.abb.2011.11.003

Jones, G., Strugnell, S.A., DeLuca, H.F.: Current understanding of the molecular actions of vitamin D. Physiological Reviews **78**(4), 1193–1231 (1998). URL `http://physrev.physiology.org/content/78/4/1193.full.pdf+html`

Kelley, C.: Interative Methods for Optimization. SIAM (1999)

Kolda, T.G., Lewis, R.M., Torczon, V.: Optimization by direct search: New perspectives on some classical and modern methods. SIAM review **45**(3), 385–482 (2003). DOI 10.1137/S003614450242889

Krishnan, K., Crouse, L., Bazar, M., Major, M., Reddy, G.: Physiologically based pharmacokinetic modeling of cyclotrimethylenetrinitramine in male rats. Journal of Applied Toxicology **29**(7), 629–637 (2009). DOI 10.1002/jat.1455

Krishnan, K., Loizou, G.D., Spendiff, M., Lipscomb, J.C., Andersen, M.E.: Quantitative Modeling in Toxicology, chap. 2: PBPK Modeling: A Primer, pp. 19–58. John Wiley & Sons, Ltd (2010)

Krogstad, H.: Digitizing graphs. MATLAB Central File Exhange (2006). URL `https://www.mathworks.com/matlabcentral/fileexchange/11640-digitizing-graphs`. Retrieved May 20, 2013

Krzyścin, J.W., Jarosławski, J., Sobolewski, P.S.: A mathematical model for seasonal variability of vitamin D due to solar radiation. Journal of Photochemistry and Photobiology B: Biology **105**, 106–112 (2011). DOI 10.1016/j.jphotobiol.2011.07.008

Leo, A., Hansch, C., Elkins, D.: Partition coefficients and their uses. Chemical reviews **71**(6), 525–616 (1971). DOI 10.1021/cr60274a001

Levy, A.B.: Stationarity and Convergence in Reduce-or-Retreat Minimization. Springer (2012)

Lips, P.: Worldwide status of vitamin D nutrition. The Journal of Steroid Biochemistry and Molecular Biology **121**(1-2), 297–300 (2010). DOI 10.1016/j.jsbmb.2010.02.021

Ljung, L., Glad, T.: On global identifiability for arbitrary model parametrizations. Automatica **30**(2), 265–276 (1994). DOI 10.1016/0005-1098(94)90029-9

Looker, A.C., Johnson, C.L., Lacher, D.A., Pfeiffer, C.M., Schleicher, R.L., Sempos, C.T., et al.: Vitamin D status: United States, 2001-2006. US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics (2011)

Manning, C.C., Schlosser, P.M., Tran, H.T.: A multicompartment liver-based pharmacokinetic model for benzene and its metabolites in mice. Bulletin of Mathematical Biology **72**(3), 507–540 (2010). DOI 10.1007/s11538-009-9459-x

McLean, K.A., McAuley, K.B., McLean, K.A., McAuley, K.B.: Mathematical modelling of chemical processes - obtaining the best model predictions and parameter estimates using identifiability and estimability procedures. The Canadian Journal of Chemical Engineering **90**(2), 351–366 (2012). DOI 10.1002/cjce.20660

Mendel, C.M.: The free hormone hypothesis distinction from the free hormone transport hypothesis. Journal of Andrology **13**(2), 107–116 (1992). DOI 10.1002/j.1939-4640.1992.tb01639.x

Miao, H., Xia, X., Perelson, A.S., Wu, H.: On identifiability of nonlinear ode models and applications in viral dynamics. SIAM review **53**(1), 3–39 (2011). DOI 10.1137/090757009

Morán-Auth, Y., Penna-Martinez, M., Shoghi, F., Ramos-Lopez, E., Badenhoop, K.: Vitamin D status and gene transcription in immune cells. The Journal of Steroid Biochemistry and Molecular Biology **136**, 83–85 (2013). DOI 10.1016/j.jsbmb.2013.02.005

Nocedal, J., Wright, S.J.: Numerical optimization. Springer Science+ Business Media (2006)

Olds, W., McKinley, A., Moore, M., Kimlin, M.: In vitro model of vitamin D3 (cholecalciferol) synthesis by UV radiation: Dose–response relationships. Journal of Photochemistry and Photobiology B: Biology **93**(2), 88–93 (2008). DOI 10.1016/j.jphotobiol.2008.07.004

Poulin, P., Haddad, S.: Advancing prediction of tissue distribution and volume of distribution of highly lipophilic compounds from a simplified tissue-composition-based model as a mechanistic animal alternative methods. Journal of Pharmaceutical Sciences **101**(6), 2250–2261 (2012). DOI 10.1002/jps.23090

Poulin, P., Krishnan, K.: A biologically-based algorithm for predicting human tissue: blood partition coefficients of organic chemicals. Human and Experimental Toxicology **14**, 273–280 (1995). DOI 10.1177/096032719501400307

Poulin, P., Theil, F.: Prediction of pharmacokinetics prior to in vivo studies. 1. mechanism-based prediction of volume of distribution. Journal of Pharmaceutical Sciences **91**(1), 129–156 (2002). DOI 10.1002/jps.10005

Prosser, D.E., Jones, G.: Enzymes involved in the activation and inactivation of vitamin D. Trends in Biochemical Sciences **29**(12), 664–673 (2004). DOI 10.1016/j.tibs.2004.10.005

Quaiser, T., Mönnigmann, M.: Systematic identifiability testing for unambiguous mechanistic modeling–application to JAK-STAT, MAP kinase, and NF-κB signaling pathway models. BMC systems biology **3**(1), 50 (2009). DOI 10.1186/1752-0509-3-50

Rajwade, M.S., Katyare, S.S., Fatterpaker, P., Sreenivasan, A.: Regulation of mitochondrial protein turnover by thyroid hormone (s). Biochemical Journal **152**(2), 379–387 (1975). URL `http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1172481/pdf/biochemj00545-0227.pdf`

Ratkowsky, D.A., Giles, D.E.: Handbook of nonlinear regression models. Marcel Dekker New York (1990)

Rattanamongkonkul, S., Sripraphot, P., Rattanakul, C.: Effect of vitamin D on bone formation and resorption: mathematical modeling approach. In: Proceedings of the 5th international conference on Applied mathematics, simulation, modelling, ASM'11, pp. 48–53. World Scientific and Engineering Academy and Society (WSEAS), Stevens Point, Wisconsin, USA (2011). URL `http://www.wseas.us/e-library/conferences/2011/Corfu/ASM/ASM-07.pdf`

Recant, L., Riggs, D.S.: Thyroid function in nephrosis. Journal of Clinical Investigation **31**(8), 789–797 (1952). DOI 10.1172/JCI102664

Robinson, C.: Dynamical systems: stability, symbolic dynamics, and chaos, vol. 28. CRC PressI Llc (1999)

Rosenstreich, S.J., Rich, C., Volwiler, W.: Deposition in and release of vitamin D3 from body fat: evidence for a storage site in the rat. Journal of Clinical Investigation **50**, 679–687 (1971). DOI 10.1172/JCI106538

Sakaki, T., Kagawa, N., Yamamoto, K., Inouye, K.: Metabolism of vitamin D3 by cytochromes P450. Frontiers in Bioscience **10**, 119–134 (2005). URL `http://www.bioscience.org/2005/v10/af/1514/fulltext.htm`

Sawada, N., Sakaki, T., Kitanaka, S., Takeyama, K.i., Kato, S., Inouye, K.: Enzymatic properties of human 25-hydroxyvitamin D3 1a-hydroxylase. European Journal of Biochemistry **265**(3), 950–956 (1999). DOI 10.1046/j.1432-1327.1999.00794.x

Sawada, N., Sakaki, T., Ohta, M., Inouye, K.: Metabolism of vitamin D(3) by human CYP27A1. Biochemical and Biophysical Research Communications **273**, 977–984 (2000). DOI 10.1006/bbrc.2000.3050

Sawyer, M.E., Evans, M.V., Wilson, C.A., Beesley, L.J., Leon, L.S., Eklund, C.R., Croom, E.L., Pegram, R.A.: Development of a human physiologically based pharmacokinetic (PBPK) model for in vitro to in vivo extrapolation of dermal permeability: A case study with lindane (In review)

Shimada, T., Yamazaki, H., Mimura, M., Inui, Y., Guengerich, F.P.: Interindividual variations in human liver cytochrome p-450 enzymes involved in the oxidation of drugs, carcinogens and toxic chemicals: studies with liver microsomes of 30 Japanese and 30 Caucasians. Journal of Pharmacology and Experimental Therapeutics **270**(1), 414–423 (1994). URL `http://jpet.aspetjournals.org/content/270/1/414.full.pdf+html`

Shinkyo, R., Sakaki, T., Kamakura, M., Ohta, M., Inouye, K.: Metabolism of vitamin D by human microsomal CYP2R1. Biochemical and Biophysical Research Communications **324**, 451–457 (2004). DOI 10.1016/j.bbrc.2004.09.073

Slob, W., Janssen, P., Van den Hof, J.: Structural identifiability of PBPK models: Practical consequences for modeling strategies and study designs. CRC Critical Reviews in Toxicology **27**(3), 261–272 (1997). DOI 10.3109/10408449709089895

Stamp, T.C.B., Round, J.M., Rowe, D.J.F., Haddad, J.G.: Plasma levels and therapeutic effect of 25-hydroxycholecalciferol in epileptic patients taking anticonvulsant drugs. BMJ: British Medical Journal **4**(5831), 9–12 (1972). DOI 10.1136/bmj.4.5831.9

Stubbs, J.R., Idiculla, A., Slusser, J., Menard, R., Quarles, L.D.: Cholecalciferol supplementation alters calcitriol-responsive monocyte proteins and decreases inflammatory cytokines in ESRD. Journal of the American Society of Nephrology **21**(2), 353–361 (2010). DOI 10.1681/ASN.2009040451

Tang, E.K., Tieu, E.W., Tuckey, R.C.: Expression of human CYP27B1 in Escherichia coli and characterization in phospholipid vesicles. FEBS Journal **279**(19), 3749–3761 (2012). DOI 10.1111/j.1742-4658.2012.08736.x

Tripkovic, L., Lambert, H., Hart, K., Smith, C.P., Bucca, G., Penson, S., Chope, G., Hyppönen, E., Berry, J., Vieth, R., et al.: Comparison of vitamin D2 and vitamin D3 supplementation in raising serum 25-hydroxyvitamin D status: a systematic review and meta-analysis. The American journal of Clinical Nutrition **95**(6), 1357–1364 (2012). DOI 10.3945/ajcn.111.031070

Vajda, S., Rabitz, H., Walter, E., Lecourtier, Y.: Qualitative and quantitative identifiability analysis of nonlinear chemical kinetic models. Chemical Engineering Communications **83**(1), 191–219 (1989). DOI 10.1080/00986448908940662

Van Etten, E., Stoffels, K., Gysemans, C., Mathieu, C., Overbergh, L.: Regulation of vitamin D homeostasis: implications for the immune system. Nutrition Reviews **66**(s2), S125–S134 (2008). DOI 10.1111/j.1753-4887.2008.00096.x

Verma, A.: An introduction to automatic differentiation. Current Science Bangalore **78**(7), 804–807 (2000). URL `http://tejas.serc.iisc.ernet.in/currsci/apr102000/tutorial1.pdf`

Vieth, R.: The pharmacology of vitamin D, including fortification strategies. In: Feldman, D., Pike, J.W., Glorieux, F.H. (eds.) Vitamin D, vol. 1, 2nd edn. Elsevier (2005)

Vitamin D Council: (2013). URL `http://www.vitamindcouncil.org/`. Accessed May 20, 2013

Wagner, C.L., Taylor, S.N., Hollis, B.W.: Does vitamin D make the world go 'round'? Breastfeeding Medicine **3**(4), 239–250 (2008). DOI 10.1089/bfm.2008.9984

Wang, Z., Lin, Y.S., Zheng, X.E., Senn, T., Hashizume, T., Scian, M., Dickmann, L.J., Nelson, S.D., Baillie, T.A., Hebert, M.F., et al.: An inducible cytochrome p450 3A4-dependent vitamin D catabolic pathway. Molecular Pharmacology **81**(4), 498–509 (2012). DOI 10.1124/mol.111.076356

White, J.H.: Vitamin D metabolism and signaling in the immune system. Reviews in Endocrine and Metabolic Disorders **13**(1), 21–29 (2012). DOI 10.1007/s11154-011-9195-z

Williams, L., Leggett, R.: Reference values for resting blood flow to organs of man. Clinical Physics and Physiological Measurement **10**, 187 (1989). DOI 10.1088/0143-0815/10/3/001

Wilson, Z., Rostami-Hodjegan, A., Burn, J., Tooley, A., Boyle, J., Ellis, S., Tucker, G.: Inter-individual variability in levels of human microsomal protein and hepatocellularity per gram of liver. British journal of clinical pharmacology **56**(4), 433–440 (2003). DOI 10.1046/j.1365-2125.2003.01881.x

Wolf, G.: The discovery of vitamin D: The contribution of adolf windaus. The Journal of Nutrition **134**(6), 1299–1302 (2004). URL `http://jn.nutrition.org/content/134/6/1299.full.pdf+html`

Xia, X., Moog, C.: Identifiability of nonlinear systems with application to HIV/AIDS models. Automatic Control, IEEE Transactions on **48**(2), 330–336 (2003). DOI 10.1109/TAC.2002.808494

Xu, Y., Hashizume, T., Shuhart, M.C., Davis, C.L., Nelson, W.L., Sakaki, T., Kalhorn, T.F., Watkins, P.B., Schuetz, E.G., Thummel, K.E.: Intestinal and hepatic CYP3A4 catalyze hydroxylation of 1alpha,25-dihydroxyvitamin D3: Implications for drug-induced osteomalacia. Molecular Pharmacology **69**(1), 56–65 (2006). DOI 10.1124/mol.105.017392

Zerwekh, J.E.: Blood biomarkers of vitamin D status. American Journal of Clinical Nutrition **87**(4), 1087S–1091S (2008). URL `http://ajcn.nutrition.org/content/87/4/1087S.full.pdf+html`

Zhu, J., Barycki, R., Chiellini, G., DeLuca, H.F.: Screening of selective inhibitors of 1α, 25-dihydroxyvitamin D3 24-hydroxylase using recombinant human enzyme expressed in escherichia coli. Biochemistry **49**(49), 10,403–10,411 (2010). DOI 10.1021/bi101488p

Zhu, J., DeLuca, H.: Vitamin D 25-hydroxylase–four decades of searching, are we there yet? Archives of Biochemistry and Biophysics **532**, 30–36 (2012). DOI 10.1016/j.abb.2012.01.013

APPENDICES

Appendix A: Model Development Supplement

This appendix contains information that is relevant to the vitamin D model development. Appendix A.1 contains tables of abbreviations that are used throughout this article andAppendix A.2 contains a description of how the initial conditions for all compartments were extrapolated.

A.1 Symbols and Abbreviations

The following symbols and abbreviations are used throughout the article.

Table A.1: Compound and enzyme abbreviations.

Name	**Abbreviation**	**Relevance**
Cholecalciferol	Vitamin D	Primary compound, dose compound
Calciferol	25(OH)D	Metabolite, has available time-course data
Calcidol	1,25(OH)2D	Metabolite, used in regulation of 25(OH)D

Continued on next page

Table A.1 – *Continued from previous page*

Name	**Abbreviation**	**Relevance**
CYP27A1	27A1	Converts vitamin D to 25(OH)D
CYP2R1	2R1	Converts vitamin D to 25(OH)D
CYP27B1	27B1	Converts 25(OH)D to 1,25(OH)2D
CYP24A1	24A1	Converts 1,25(OH)2D to further metabolites
CYP27A1b	27A1b	Converts 25(OH)D to 1,25(OH)2D
CYP24A1b	24A1b	Converts 25(OH)D to further metabolites

Table A.2: Compartmental abbreviations

Symbol	**Description**	**Notes**
L	Liver	Main metabolizing organ
K	Kidney	Secondary metabolizing organ
GI	GI Tract	Stomach and small intestines, entry from oral supplementation
S	Skin	Entry from solar synthesis
RP	Rapidly Perfused Tissue	Heart, brain, lungs, pancreas, and thyroid
SP	Slowly Perfused Tissue	Extrapolated from other compartments
Adi	Adipose	Adjusted volume using study BMI values
Spl	Spleen	Unique contribution to hepatic circulation
Art	Arterial Blood	25% total blood volume
Ven	Venous Blood	75% total blood volume
Λ	$\{K, S, RP, SP, Adi\}$	Used in equations for venous compartment

Table A.3: Primary Symbols. Units of symbols are given in parentheses.

Symbol	Description
C_j^i	Concentration of chemical i in compartment j, namely $C_j^i = \frac{A_j^i}{V_j \cdot P_j^i}$ (ng/L)
Q_j	Blood flow in tissue j (L)
Q_{hep}	Hepatic blood flow, where $Q_{hep} = Q_L + Q_{GI} + Q_{Spl}$ (L)
V_{max}^i	Maximum rate of metabolism for enzyme i (ng/hr)
K_m^i	Concentration at half-saturation for enzyme i (ng/L)
K_{abs}	Rate of absorption from the GI lumen to the GI tract (ng/hr)
A_j^i	Amount of chemical i in compartment j (ng)
AA^i	Amount of chemical i in arterial compartment (ng)
AV^i	Amount of chemical i in venous compartment (ng)
CA^i	Concentration of chemical i in arterial compartment, $CA^i = \frac{AA^i}{V_{Art}}$ (ng/L)
CV^i	Concentration of chemical i in venous compartment, $CV^i = \frac{AV^i}{V_{Ven}}$ (ng/L)
V_j	Volume of compartment j (L)
P_j^i	Tissue j:blood partition coefficient for chemical i (unitless)
BW	Body weight (kg)

A.2 Prediction of Initial Conditions

In typical PBPK models, the dosing compound is assumed to be novel, meaning that the body has little to no prior exposure to the compound. This is not the case for the vitamin D cascade. Not only is vitamin D present in many fortified foods such as milk and cereals,

Table A.4: Kinetic parameter ranges and literature sources. *Units are pmol/min/mg p450.

Enzyme	$\mathbf{K_m}$ (μM)	$\mathbf{V_{max}}$ (mol/min/mol p450)	**Reference**
CYP27A1	3.2 ± 0.5	0.27 ± 0.03	(Sakaki et al., 2005; Shinkyo et al., 2004)
CYP2R1	0.45 ± 0.16	1.2 ± 0.05	(Shinkyo et al., 2004)
CYP27B1	2.7 ± 0.07	$3.9 \pm 1.6^*$	(Sawada et al., 1999)
CYP24A1	0.17 ± 0.05	0.105 ± 0.012	(Zhu et al., 2010)
CYP27A1b	2.8 ± 0.4	0.021 ± 0.002	(Sawada et al., 2000)
CYP24A1b	0.33 ± 0.04	0.12 ± 0.012	(Zhu et al., 2010; Zhu and DeLuca, 2012)

it is also internally synthesized in response to sunlight exposure. Whether ingested or synthesized, vitamin D and the primary metabolites are present in all compartments that we consider in this model prior to the start of the experimental.

This lack of novelty requires that the initial conditions of vitamin D and its metabolites be estimated *a priori*. By initializing each compartment before to the start of the experiment, the model can avoid an equilibrating process that would otherwise skew model predictions. Our method also allows us to tailor the initial amount to the dosing level based on the first data point. Through the design of the experiment, we have an approximation of the daily intake of vitamin D through fortified foods as well as an average amount of 25(OH)D in venous blood for each dose level. We will use these values in conjunction with estimates of vitamin D and 1,25(OH)2D serum levels to predict the associated levels based on the initial 25(OH)D concentration.

Method

Three separate sources are used for obtaining the initial concentrations of each metabolite in the venous compartment. We chose sources that allowed us to compare the ratio of two metabolites, one of which is 25(OH)D, so that we could scale the initial amounts to the dataset from (Fisk et al., 2012). To determine vitamin D concentrations, we extrapolated the ratio of the concentration of vitamin D to 25(OH)D from baseline data in (Heaney et al., 2003); this ratio is approximately equal to 6%. 1,25(OH)2D concentrations were taken from (Chun et al., 2012) and verified against data in (Bikle et al., 1984); the ratio of 1,25(OH)D to 25(OH)D is approximately equal to 0.2%. Finally, dose-dependent concentrations of 25(OH)D were taken from (Fisk et al., 2012) and were used to scale the initial amounts of the other compounds. These amounts can be found in Table A.5.

Table A.5: Initial (total and free) concentration of metabolites in the venous compartment.

		Dose		
	Amount	Zero	5μg	10μg
Vitamin D	Total (nM)	1.93	2.03	2.06
	Free (pM)	1.4	1.4	1.4
25(OH)D	Total (nM)	32.11	33.89	34.35
	Free (pM)	12.8	13.6	13.7
1,25(OH)2D	Total (nM)	0.1	0.1	0.1
	Free (pM)	0.4	0.4	0.4

Table A.5 also includes rows that give the free amount of each compound. These

amounts are constructed using the Free Hormone Hypothesis (FHH), as discussed in Section 2.2.1. Note in previous discussions, we gave reasons why the FHH was not valid for vitamin D; these arguments apply strictly to orally-absorbed vitamin D, not vitamin D produced by the body. For internally-produced vitamin D, the FHH does, in fact, hold. Therefore, our scaling factor of 0.07% is justified in Table A.5 for vitamin D.

Once we have initial amounts in the venous compartment for each metabolite, we are interested in finding the amounts in all other compartments. To do this, we will assume a closed system PBPK model for each metabolite. These models are constructed by looking at the compartments for each metabolite and removing any terms that allow entrance or exit of compounds from the system, such as the kinetic-related terms. This creates three isolated systems, one for each compound.

We can now consider each of these systems separately. As an example, let us consider the case for 25(OH)D. We have extrapolated the free amount of 25(OH)D in the venous compartment under steady-state conditions. Therefore, we need to find the amounts in the other compartments that will maintain this steady state amount. Under these steady-state conditions, neither the adipose partition coefficients nor the amount of p450 enzymes available (the two adaptions to the original model) will change. Let the concentration of 25(OH)D in the venous compartment be represented by C. To calculate the concentrations in the other compartments, we solve for a positive scalar β such that with the initial conditions βC in the venous compartment and zero elsewhere, the 25(OH)D system will reach an equilibrium such that the concentration in the venous compartment equals C and the remaining compartments are also at equilibrium.

The other cases follow in the same manner, differing only in the scalar β that yields the equilibrium values.

Appendix B: Approximations and Optimizations

A major portion of data fitting requires calculating the Fisher Information Matrix. To do this, we need to calculate numerous partial derivatives. This section discusses techniques for calculating this matrix, as well as optimization routines that are used in the process of finding a best fit to model data.

B.1 Calculation of Sensitivity Coefficients

Recall the setup of sensitivity-based identifiability analysis from Section 4.1.3:

$$\frac{d\mathbf{x}(t)}{dt} = f(t, \mathbf{x}; \mathbf{q}) \tag{B.1}$$

where $\mathbf{x} \in \mathbb{R}^n$ denotes the states and $\mathbf{q} \in \mathbb{R}^m$ denotes the parameters. To determine the effects of parameters $\mathbf{q}$ on states $\mathbf{x}$, we can differentiate both sides of the equation with

respect to $\mathbf{q}$ to obtain a system of $n + mn$ differential equations:

$$\frac{d\mathbf{x}(t)}{dt} = f(t, \mathbf{x}; \mathbf{q}) \tag{B.2}$$

$$\frac{d}{dt}\frac{\partial \mathbf{x}}{\partial \mathbf{q}} = \frac{\partial f}{\partial \mathbf{x}}\frac{\partial \mathbf{x}}{\partial \mathbf{q}} + \frac{\partial f}{\partial \mathbf{q}}. \tag{B.3}$$

Note the mn sensitivity equations in Equation (B.3) are linear, matrix differential equations in terms of $\frac{\partial \mathbf{x}}{\partial \mathbf{q}}$. However, $\frac{\partial f}{\partial \mathbf{x}}$ and $\frac{\partial f}{\partial \mathbf{q}}$ must be computed, either through finite differentiation or similar manual techniques or through the use of Automatic Differentiation (AD).

B.1.1 Finite Difference Method

In many practical applications, a scalar-valued function f is not given by a formula, but rather is the output from a computational or experimental procedure. This generally implies that the derivative f' is not readily available, and must be approximated when required. Consider the case where $x \in \mathbb{R}^1$. Let h be a small, positive number. Then the Taylor series of order n of a function f is given by

$$f(x + h) = f(x) + hf'(x) + \frac{h^2}{2}f''(x) + \cdots + \frac{h^n}{n!}f^{(n)}(x). \tag{B.4}$$

The derivative $f'(x)$ can be written as

$$f'(x) = \frac{f(x + h) - f(x)}{h} + \mathscr{O}(h). \tag{B.5}$$

Equation (B.5) is referred to as the *forward finite difference*. Alternately, utilizing the Taylor series expansion of $f(x-h)$ yields

$$f'(x) = \frac{f(x) - f(x-h)}{h} + \mathscr{O}(h), \tag{B.6}$$

also known as the *backward finite difference*. Combining these two equations yields the *central finite difference formula*:

$$f'(x) = \frac{f(x+h) - f(x-h)}{2h} + \mathscr{O}(h^2). \tag{B.7}$$

As biological systems often contain more than one dimension, consider the multivariable case where $F : \mathbb{R}^n \to \mathbb{R}^m$. The derivative of F at $\mathbf{x}$ is called the Jacobian (J) of F at $\mathbf{x}$, where

$$F'(\mathbf{x})_{ij} = \frac{\partial f_i}{\partial \mathbf{x}_j}(\mathbf{x}), \quad F'(\mathbf{x}) = J(\mathbf{x}). \tag{B.8}$$

If $J(x)$ is not readily available, it can be approximated in a similar manner as in the one-dimension case by using finite difference approximations. Here, we use the central finite difference. The jth column of $J(\mathbf{x})$ is approximated by

$$A_j = \frac{F(\mathbf{x} + he_j) - F(\mathbf{x} - he_j)}{2h}, \tag{B.9}$$

where e_j is the jth unit vector.

A reasonable question to ask is how to calculate the finite-difference step sizes in practice. As most systems of equations are solved on computers, it is important to consider the differences between finite-precision and exact arithmetic. Exact arithmetic suggests choosing h as small as possible; however, finite-arithmetic may introduce errors through

inaccuracies in function values and cancellation errors. If h is too small, $\mathbf{x} + he_j$ may be so close to $\mathbf{x}$ that $F(\mathbf{x} + he_j) \cong F(\mathbf{x})$ and A_j may have very few good digits. Therefore, h must be chosen to balance these conflicting issues. Let the relative error in computing $F(\mathbf{x})$ be denoted as η. In an ideal case (considering the forward difference), we would like

$$\frac{\|F(\mathbf{x} + he_j) - F(\mathbf{x})\|}{\|F(\mathbf{x})\|} \leq \sqrt{\eta}, \quad j = 1, \ldots, n.$$

Dennis and Schnabel (1996) suggest that a reasonable way to accomplish this is to perturb each component $\mathbf{x}_j$ by its own step size $h_j = \sqrt{\eta}\ \mathbf{x}_j$, which constitutes a relative change of $\sqrt{\eta}$ in $\mathbf{x}_j$. In the biological models we are considering, $F(\mathbf{x})$ is given by relatively simple formulas; thus, we can assume that η is on the order of machine precision ε_{mach}, i.e. $\eta \cong \varepsilon_{mach}$. Therefore, Equation (B.9) can be rewritten as

$$A_j = \frac{F(\mathbf{x} + h_j e_j) - F(\mathbf{x})}{h_j}, \tag{B.10}$$

with $h_j = \sqrt{\eta}\ \mathbf{x}_j$. Thus, the forward difference approximation perturbs each component in a relative fashion, and the choices of h_j minimize the round-off and machine error in calculations. In the case of the central difference approximation, we utilize $h = \sqrt[3]{\eta}\ \mathbf{x}_j$.

B.1.2 Automatic Differentiation

Automatic Differentiation (AD) is a chain-rule-based technique for evaluating derivatives with respect to input variables (Griewank, 1989; Verma, 2000). This can be accomplished by noting that all computer programs use a finite set of elementary operations (e.g., trigonometric functions, addition, and multiplication are examples) to define a mathematical function through composition of these elementary operations. As the partial

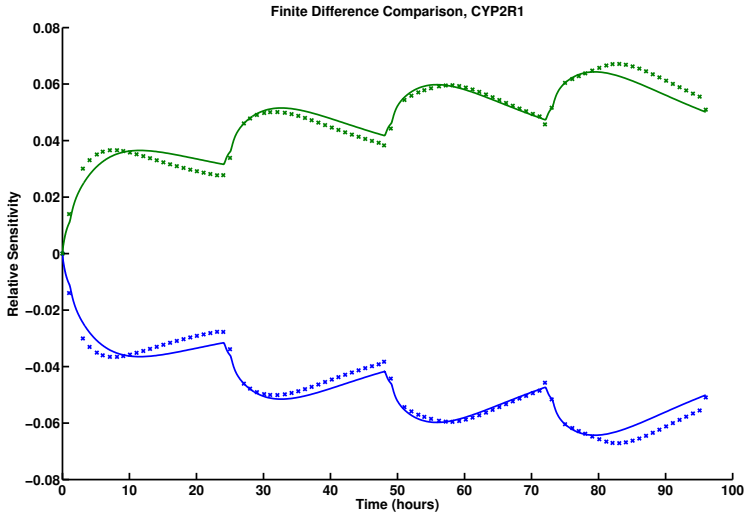
derivatives of these elementary functions are known, the total derivative is computed through applying the chain rule to all functional compositions. The result from AD is numerically accurate to floating point, thus preserving accuracy unlike finite difference methods (Carmichael et al., 1997).

B.1.3 Comparison of Calculation Techniques

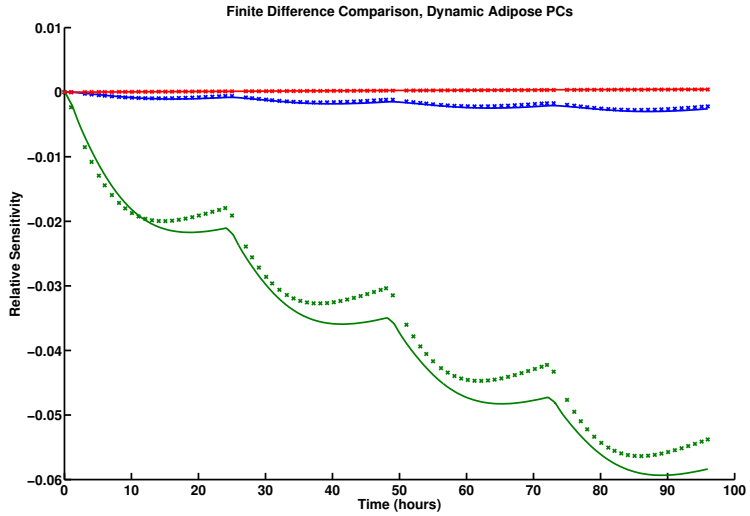
In the process of solving the sensitivity equations in Equation (B.3), we are required to calculate several partial derivatives. We have presented two techniques to accomplish this: a finite difference method and Automatic Differentiation. In Figure B.1, we compare the sensitivity coefficients obtained by using each method and show that the central finite difference approximation fits the results from Automatic Differentiation nicely. To illustrate the tightness of the fit, we show results over a time span of 96 hours (4 days). The 'x' markers indicate the finite difference approximation and the solid lines indicate calculation of the sensitivity coefficients using AD techniques.

B.2 Optimization Techniques

After determining the identifiable and estimable parameters, we utilized several techniques to optimize the parameters for a suitable model fit. These methods included direct search methods and gradient-based methods. Direct search methods are often a good first choice for PBPK models, as derivatives of the model are not needed. After the direct search method finds parameters in a lower cost region, we switch to a gradient-based method to speed up convergence.



(a) Finite difference comparison for CYP2R1



(b) Finite difference comparison for dynamic adipose PC

Figure B.1: Comparison of sensitivity calculations.

B.2.1 Constraint Transformations

Many optimization algorithms in MATLAB do not directly admit constraints on variables. An alternate option is to change the optimization algorithm to one that does admit constraints; however, we are interested in utilizing the speed and complexity of algorithms such as the Nelder-Mead direct search method or the Levenberg-Marquardt gradient-based search algorithm.

The first attempt that we made on constraining variables was to have the cost function return a value of "NaN" (Not-a-Number) whenever the parameter choice stepped out of bounds. The NaN value was chosen over an arbitrary high cost as the optimizer function handles NaNs in a slightly different manner. However, we found that with a NaN, the optimizer would over-penalize the step direction.

A second approach is to transform the variables such that the bounds vanish, but the relationship between the variable and the bounds are preserved. The use of a a scaled periodic function will accomplish this. We draw inspiration from (D'Errico, 2006) in our following explanation.

We will consider the case where all parameters are bounded above and below by a finite value. Let $\mathbf{x}_{init} = [x_1, x_2, \ldots, x_n]^T$ be the vector of parameters, and L and U be the corresponding lower and upper bounds of $\mathbf{x}_{init}$. Let $\mathbf{u}$ denote the transformed parameters. We first scale each parameter by the magnitude of the bounding region:

$$\hat{\mathbf{x}}(i) = 2\left(\frac{\mathbf{x}_{init}(i) - L(i)}{U(i) - L(i)}\right) - 1. \tag{B.11}$$

We then take this scaled parameter $\hat{\mathbf{x}}$ and apply the arcsine function:

$$\mathbf{u}(i) = 2\pi + \arcsin\big(\max(-1, \min(1, \hat{\mathbf{x}}(i)))\big). \tag{B.12}$$

The addition of 2π is included to shift the transformed variables away from the origin, which may cause problems in the Nelder-Mead search algorithm. The use of the combined max/min functions is to remedy any floating point issues. The arcsine function has an unbounded range, and thus the original bounded parameters are now scaled and transformed into unbounded values.

Once Equations (B.11) and (B.12) are used to transform the original parameters, $\mathbf{u}$ is fed into the optimizing function. As calculations in the cost function rely on the original parameters, we must undo the transformation just before the parameters are passed to the cost function. The backward-transformation is simple as well: Let $\hat{\mathbf{u}}$ denote an intermediate step, as before with $\hat{\mathbf{x}}$.

$$\hat{\mathbf{u}}(i) = \left(\frac{\sin(u(i)) + 1}{2}\right) \times (U(i) - L(i)) + L(i). \tag{B.13}$$

To compensate for potential floating point issues, we again apply a combination of max/min functions to Equation (B.13):

$$\mathbf{x}_{new}(i) = \max\big(L(i), \min(U(i), \hat{\mathbf{u}}(i))\big). \tag{B.14}$$

The values in $\mathbf{x}_{new}$ correspond to choices for new parameter values. The cost function returns a value to the optimizer for the cost of the transformed parameter selection, at which point the algorithm selects for new parameters, as necessary.

The benefits of this transformation is that introduction of NaN values is not necessary,

which allows for more consistency and speed within the optimizing routine. We directly use the function `fminsearchbnd` created by (D'Errico, 2006) when utilizing the Nelder-Mead search algorithm, and modifications of this function for other optimization algorithms.

B.2.2 Direct Search Methods

Direct search methods are deterministic sampling methods, meaning they only require function values to compute components of the algorithm. These methods do not need information about the derivatives of f to determine the minimizer of the function; instead, direct search methods utilize sequential comparison of trial solutions with the "best" solution obtained up to that time point. This is coupled with a strategy to decide what the next trial solution will be. A more formal definition is given in (Kolda et al., 2003) and contains the following requirements:

1. An order relation $\prec$ exists between any two points $\mathbf{x}$ and $\mathbf{y}$ such that $\mathbf{x} \prec \mathbf{y}$ if $f(\mathbf{x}) < f(\mathbf{y})$ for some function f to be minimized. This means that $\mathbf{x}$ is a "better" point than $\mathbf{y}$ because it yields a lower function value.

2. At any iteration, only a finite number of new iterates exist and these possibilities can be determined in advance.

For this model, we used two direct search methods: the Nelder-Mead (NM) search algorithm and the Compass Search algorithm. Both algorithms are described in the following subsections. We chose to run the optimization routine twice, once with each direct search algorithm. The resulting parameter values were similar, but the NM algorithm, utilized with transformation algorithm described in Appendix B.2.1, took significantly less time to find a local minimum than the compass search method. Therefore, the

final optimization routine that we chose uses the NM search algorithm with constraint transformations.

An addition side benefit to using the NM search algorithm over the compass search algorithm is that the MATLAB function for NM is available in the standard distribution package, whereas the compass search algorithm requires the Global Optimization toolbox.

Nelder-Mead Search Algorithm

A significant portion of the optimization utilized the Nelder-Mead (NM) search algorithm to find optimal parameter values. Consider a set of vertices $\{x_j\}$ for $j = 1, 2, \ldots, N+1$ for $x_j \in \mathbb{R}^N$ and a function to be minimized, f. These vertices are sorted according to the objective function values

$$f(x_1) \leq f(x_2) \leq \ldots \leq f(x_{N+1}). \tag{B.15}$$

The NM algorithm attempts to replace the "worst" point x_{N+1} (the point with the highest function value, $f(x_{N+1})$) with a new point that has a lower function value with the form

$$x(\alpha) = (1+\alpha)\bar{x} - \alpha x_{N+1}. \tag{B.16}$$

The simplex in $\mathbb{R}^N$ is constructed of $\{x_j\}_{j=1}^N$ and the centroid, $\bar{x}$ is computed:

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{B.17}$$

The value of α is selected such that

$$-1 < \alpha_{ic} < 0 < \alpha_{oc} < \alpha_r < \alpha_e, \tag{B.18}$$

where α_k denotes an inner contraction ic, an outer contraction oc, a reflection r, or an expansion e of the worst point. Standard values for α are

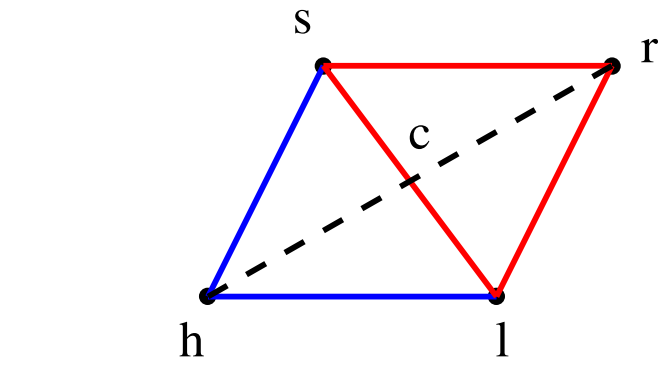
$$\{\alpha_r, \alpha_e, \alpha_{oc}, \alpha_{ic}\} = \{1, 2, 1/2, -1/2\}$$

It is possible that none of these new points results in a function value less than $f(x_{N+1})$. In this case, the algorithm shrinks the simplex toward the "best" point, $f(x_1)$ by replacing other points with

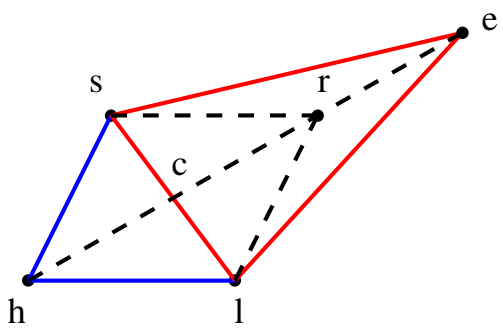
$$\bar{x}_i = \frac{x_i + x_1}{2}. \tag{B.19}$$

The algorithm then restarts by sorting the points and finding the best and worst vertices. To illustrate this method, consider the case when $N = 2$. The parameter space is $\mathbb{R}^2$ and the simplex appears as a triangle in the plane. Let h, s, and l denote these vertices. Figure B.2 demonstrates the various directions that the algorithm can take the worst point h. Here, r denotes the reflection point, and is shown is each image for a sense of scale. The current simplex is shown in blue, various sight-lines in dashed black, and the new working simplex in red. Figure B.2a demonstrates the effect of reflecting h across the centroid to r. Figure B.2b shows the simplex extending to the expansion point e, keeping the working simplex as large as possible to prevent premature termination of iterations (this is especially helpful for non-smooth functions). Figures B.2c and B.2d illustrate the effects of contraction of the simplex, whereas Figure B.2e shows the effect of shrinking the simplex toward the best point, l due to lack of improvement in the function value.

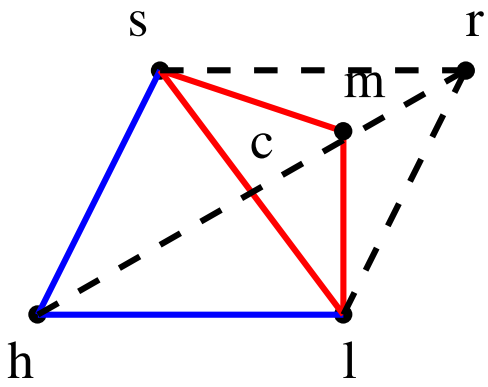
There are various stopping criteria for the NM algorithm including: iteration count, magnitude of difference between lowest and highest function values, and simplex size. For a detailed treatment of the Nelder-Mead algorithm and other similar algorithms, see (Kelley, 1999; Levy, 2012).



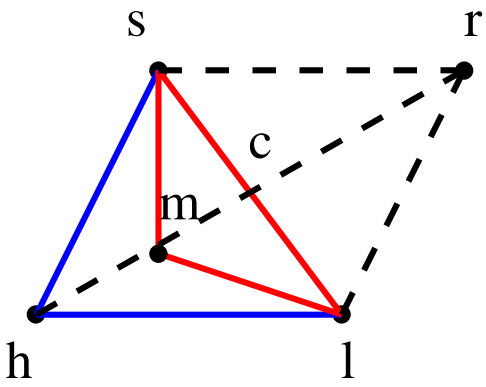
(a) Reflection



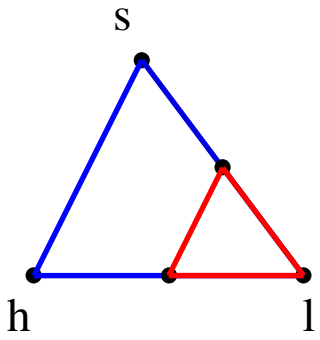
(b) Expansion



(c) Outer contraction



(d) Inner contraction



(e) Shrink

Figure B.2: Various transformations of the working simplex.

Compass Search Method

Another direct search method that we attempted to utilize in this model development is the *Compass Search Method*, implemented using the function `patternsearch` in the Matlab global optimization toolbox. The compass search method chooses points p_i equidistant from the starting point s in each of the coordinate directions; the function f is evaluated at s and p_i. If $f(p_j) < f(s)$ for some j, then move the starting point to p_j and restart. Otherwise, contract the step-length of the search directions and restart. If the step-length is smaller than a predetermined minimum step-length, then terminate the search. Figure B.3 demonstrates these techniques.

Using the direct search method guarantees an eventual convergence to a minimum. To see this, consider the two-dimensional case. Figure B.3 indicates that at each iteration, at least one of the four directions is a descent direction, unless the starting point is fixed. Extrapolation to an n-dimensional case is easy: given any $\mathbf{x} \in \mathbb{R}^n$ such that $\nabla f(\mathbf{x}) \neq 0$, at least one coordinate direction is in a descent direction. Searching in multiple directions implicitly also ensures a descent direction, perhaps steeper than just a single direction. The remainder of the proof of convergence is given in (Kolda et al., 2003); the interested reader is also referred to this reference for further discussion on modifications of the compass search method as well as additional direct search optimization methods.

B.2.3 Gradient-Based Methods

Direct search methods are typically good at getting into a region of a local minimum, but tend to stall as the step-sizes of the optimizer decrease. To combat this stalling, we can turn to gradient-based methods which utilize the derivative (or an approximation of the derivative) to find the minimum. There are several different methods that are

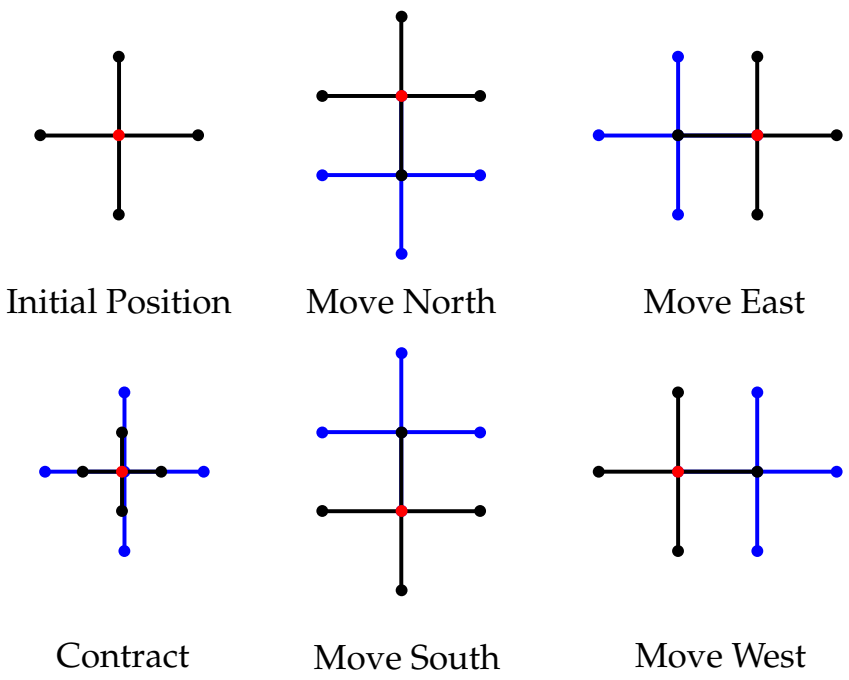


Figure B.3: Compass Search step directions. The blue compass indicates the location of the previous iteration.

available for gradient-based optimizers; for the purposes of this model, we will use the Levenberg-Marquardt Algorithm.

Levenberg-Marquardt Algorithm

The Levenberg-Marquardt Algorithm (LMA) can be seen as an adaptation of the Gauss-Newton algorithm (a modification of Newton's Method) and the method of gradient descent. Rather than using a line search as is the case with the Gauss-Newton algorithm, LMA utilizes a trust-region strategy. Briefly speaking, a trust region is a small subset of the region of the objective function to be optimized that is approximated using a model function (typically quadratic). More formally, we have the following definition:

Definition B.2.1 (Trust Region)**.** *Let Δ be the radius of the ball about x_c in which the quadratic model*

$$m_c(x) = f(x_c) + \nabla f(x_c)^T(x - x_c) + (x - x_c)^T H_c(x - x_c)/2$$

can be trusted to accurately represent the function f. Δ is called the trust region radius *and the ball*

$$\mathcal{T}(\Delta) = \{x | \, \|x - x_c\| \leq \Delta\}$$

is called the trust region*. (Kelley, 1999)*

Suppose

$$f(x) = \frac{1}{2}\sum_{j=1}^{m} r_j^2(x),$$

where $r_k = (y_k - f(x_k))$ is the residual for point k. For the LMA, we are solving the subproblem

$$\min_p \frac{1}{2}\|J_k p + r_k\|_2^2 \quad \text{s.t. } p \in \mathcal{T}(\Delta). \tag{B.20}$$

As stated in (Nocedal and Wright, 2006), this is, in effect, choosing the model function as

$$m_k(p) = \frac{1}{2}\|r_k\|^2 + p^T J_k^T r_k + \frac{1}{2} p^T J_k^T J_k p.$$

Solving the trust-region subproblem Equation (B.20) amounts to determining $\lambda > 0$ such that $\|p\| = \Delta$ and

$$(J^T J + \lambda I)p = -J^T r. \tag{B.21}$$

For large values of λ, the direction p tends toward the steepest descent direction, although the magnitude of the direction decreases. Thus, λ can be controlled to always ensure descent.

An algorithm for implementation of the LMA is found in Nocedal and Wright (2006) and we use the MATLAB function `lsqnonlin` as our optimizer. Note that we do use a wrapper function to enable the use of bounds; the LMA does not directly admit optimization of bounded variables, so we use the transformation algorithm as described in Appendix B.2.1.

Appendix C: Previous Work

PBPK models can vary widely both in complexity and in purpose. This appendix contains summaries of two summer projects that were conducted in conjunction with undergraduates. The first section details a PBPK model that was created with the intention of exploring different metabolic hypotheses; the second section discusses work that utilizes PBPK models for *in vitro* to *in vivo* extrapolation.

C.1 REU 2011: Metabolism of Bromochloromethane

The following is a summary of work published in (Cuello et al., 2012).

Bromochloromethane (BCM), a volatile water disinfection byproduct of chlorination, is suggested to have dual-component metabolic pathway: a saturable Michaelis-Menten P450 enzymatic hyperbola superimposed with a linear, glutathione transferase (GST) pathway. However, recent work with a similar compound, dichloromethane, has suggested that a single enzymatic pathway–utilizing an enzyme with a dual-binding site–may

be responsible for the observed kinetics. The purpose of the summer project was to demonstrate, mathematically, that a two-binding site metabolic pathway for BCM is possible.

A flow-limited PBPK model was constructed to describe a previously published experiment. In this experiment, bolus doses of BCM were administered into a chamber in which rodents were placed. Time-course data of levels of BCM in the chamber were taken at select intervals. This experiment was repeated with five increasing doses to try to capture the entire spread of the kinetic behavior.

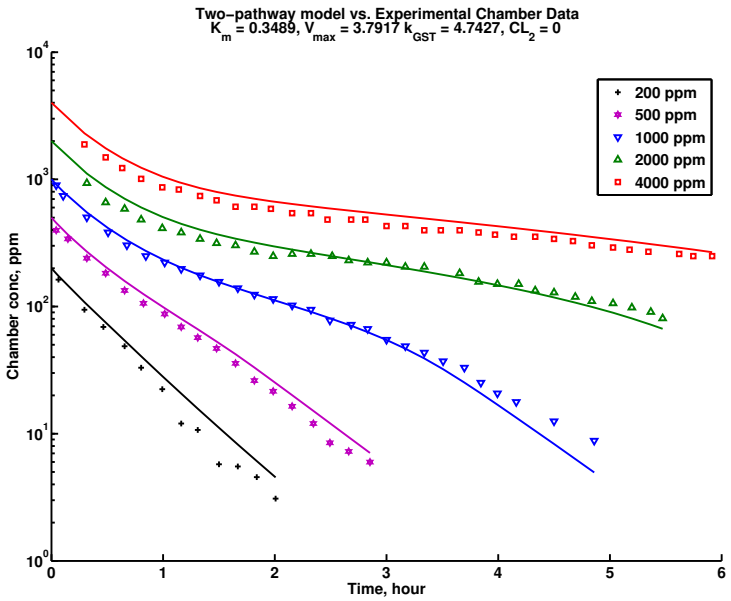
The kinetic behavior of the enzyme pathway was described using one of two models. For the accepted two-pathway model, kinetics in the liver, Met_{liv} and the kidney, Met_{kid}, were described using Equation (C.1).

$$\begin{aligned} Met_{liv} &= 0.948 \left(\frac{V_{max}[C_{liv}]}{K_m + [C_{liv}]} + k_{GST}[C_{liv}]V_{liv} \right), \\ Met_{kid} &= 0.052 \left(\frac{V_{max}[C_{kid}]}{K_m + [C_{kid}]} + k_{GST}[C_{kid}]V_{kid} \right), \end{aligned} \tag{C.1}$$

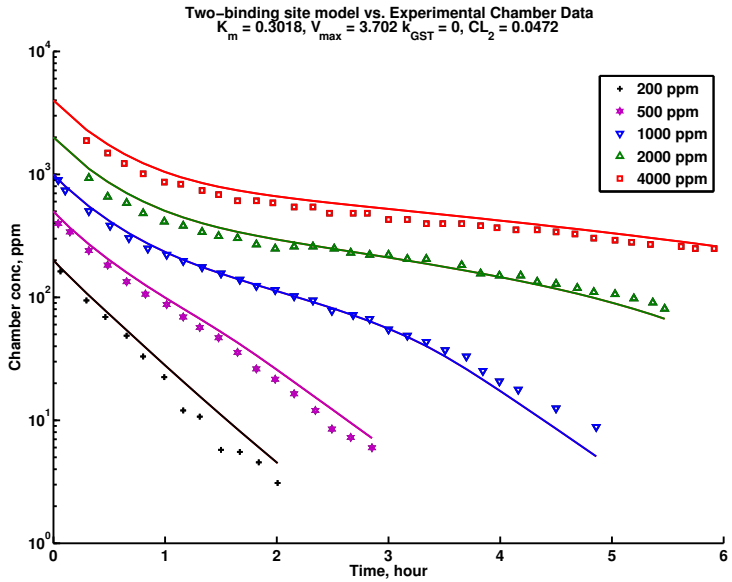
in contrast to the equations describing the two-binding site kinetics (Equation (C.2)):

$$\begin{aligned} Met_{liv} &= 0.948 \left(\frac{V_{max}[C_{liv}] + CL_2[C_{liv}]^2}{K_m + [C_{liv}]} \right), \\ Met_{kid} &= 0.052 \left(\frac{V_{max}[C_{kid}] + CL_2[C_{kid}]^2}{K_m + [C_{kid}]} \right), \end{aligned} \tag{C.2}$$

Two PBPK models, each containing one of the metabolic hypotheses, were constructed and parameters were optimized to fit available data. Figure C.1 shows the final fit. The models are virtually indistinguishable based on the fit to data; this suggests that both metabolic hypotheses are mathematically plausible.



(a) Two-pathway metabolic hypothesis



(b) Two-site metabolic hypothesis

Figure C.1: Comparison of model fit to the two-pathway (left) and two-site (right) metabolic hypotheses. Figures adapted from (Cuello et al., 2012).

In addition to optimizing for model fit, we also investigated the sensitivity of each kinetic coefficient in the chamber as well as the liver compartments. Figure C.2 shows the time-course sensitivity of the kinetic parameters in the two-binding site case. As

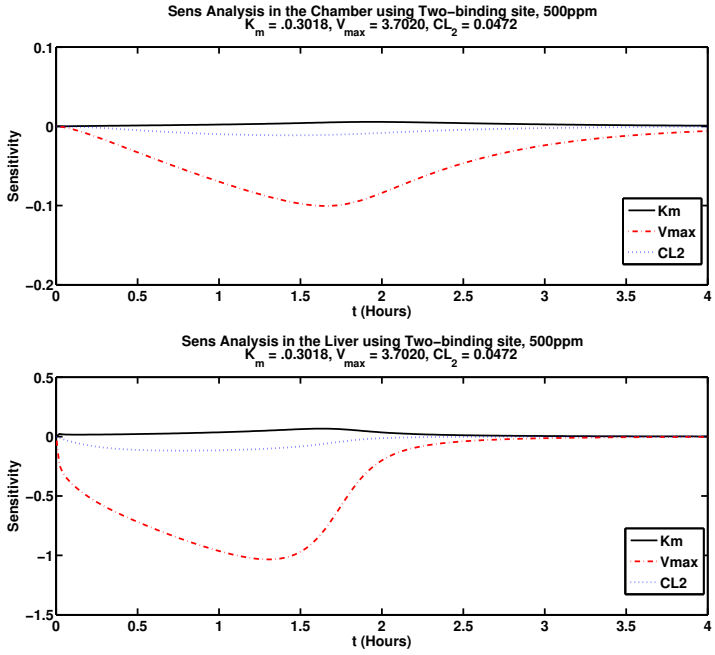


Figure C.2: Time-course sensitivities for fixed parameter values. Adapted from (Cuello et al., 2012)

sensitivity analysis is a local result, I developed a novel method to show the change in the sensitivity topography as one parameter is allowed to change. All other parameters are held constant, and the time-course sensitivity curves are plotted against the varying parameter. An example is shown in Figure C.3, where V_{max} is allowed to vary in the two-pathway model. Two graphs are presented, comparing the sensitivity in the chamber versus in the liver. This three-dimensional method allows for a better understanding of how the sensitivities change if an estimate for a parameter is inaccurate; this can allow for

better experimental design to capture an appropriate region of interest. For example, if the true value of V_{max} is larger than calculated, it will be most sensitive in the first hours of the experiment; conversely, a lower true V_{max} requires longer experimental run time, but in both cases, three hours is sufficient to capture the most sensitive behavior of V_{max}. Similar experimental insights can be extrapolated by viewing the other parameters.

C.2 REU 2012: Dermal Exposure of Lindane

The following is a summary of work found in (Sawyer et al., in review).

Physiologically-based pharmacokinetic models can be used for extrapolating results within species and between species; PBPK models can also be used to extrapolate results between types of experiments, such as from *in vitro* experiments to *in vivo* experiments. Also known as IVIVE, *in vitro/in vivo* extrapolation has significant implications in determining potential risks and benefits of introducing a drug into a population. *In vitro* experimentation is often less costly and does not carry the same ethical risks as *in vivo* testing; however, drugs may interact with tissues differently in a laboratory setting versus in actual implementation.

A major portion of PBPK modeling of environmental compounds has been devoted to volatile compounds, and less to dermal exposure of semi-volatile or nonvolatile compounds. A specific issue for dermal models is the increased complexity of the skin tissue as a chemical passes through each barrier level. The amount of complexity needed to describe the compartment may be related to the lipophilicity of the compound applied in addition to the vehicle of application.

Lindane is an insecticide used in treatment of lice and scabies in humans, and although no longer used as an agricultural pesticide, is still approved as a second-tier approach

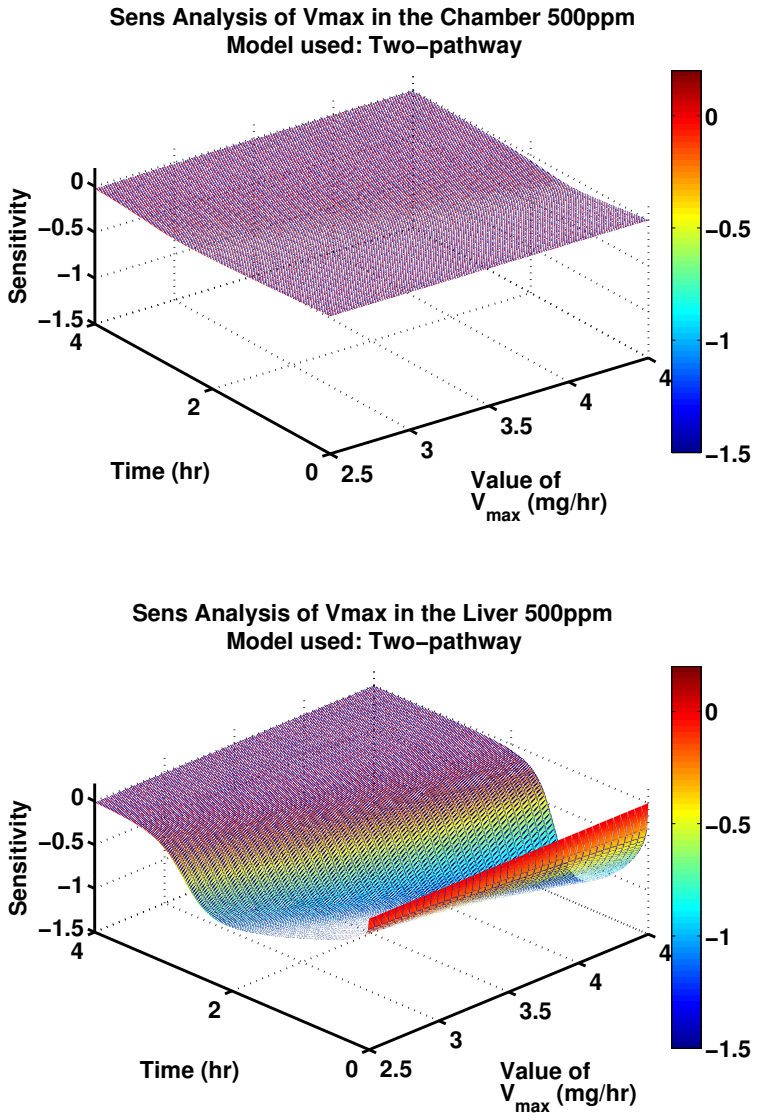


Figure C.3: 3D sensitivity analysis allowing V_{max} to vary in the two-pathway model. Adapted from (Cuello et al., 2012).

for the treatment of lice and scabies. Because of its use as a pharmaceutical, the human pharmacokinetics of the metabolism of lindane have been extensively studied in both *in*

vitro and *in vivo* dermal exposures. We utilize the results of these studies to investigate the effects of dermal compartmental construction on IVIVE of lindane.

To begin the extrapolation process, a PBPK model for the *in vitro* experimental design was created. As the dermal compartment structure–represented as either a single homogenous compartment or with a parallel follicular compartment–may influence the effects of absorption of lindane, we chose to compare each dermal model separately to the available *in vitro* data to tease out any important characteristically differences in the dermal structure. The resulting fit is shown in Figure C.4.

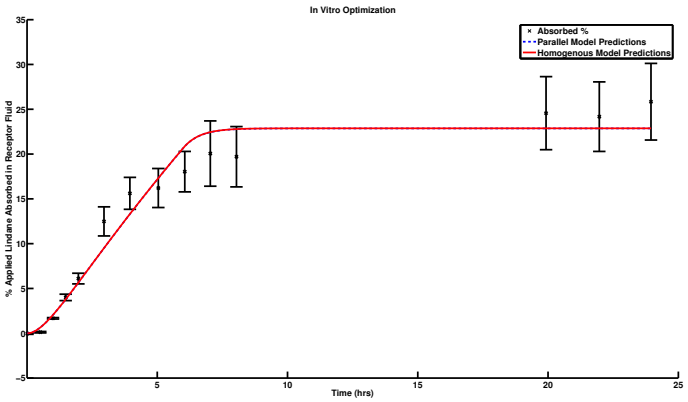


Figure C.4: *In vitro* optimization results for lindane. Adapted from (Sawyer et al., in review).

For the *in vitro* model, both the parallel and the homogenous compartments fit the available data. Utilizing several parameters that had been found in the course of optimizing for the *in vitro* model, we optimized for remaining values to construct the corresponding *in vivo* PBPK model, shown in Figure C.5.

The *in vivo* model, in contrast to the *in vitro* model, demonstrated different behavior depending on the construction of the dermal compartment. The homogenous model

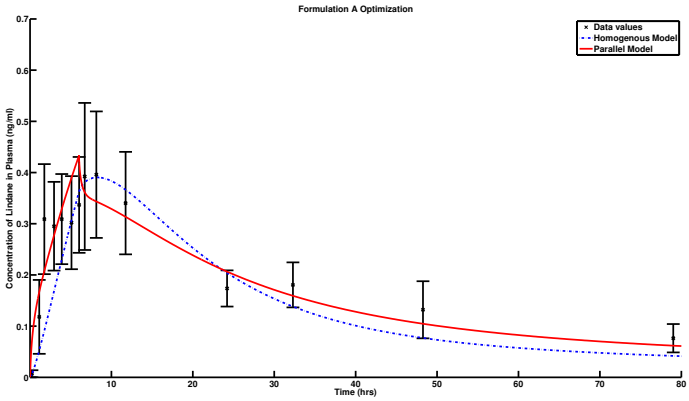


Figure C.5: *In vivo* optimization results for lindane. Adapted from (Sawyer et al., in review).

neglected to capture any of the initial absorption of the chemical, nor did it predict the slow decrease of chemical in the blood at later time points. In contrast, utilizing a parallel compartment model–where follicles act as a sort of shunt to the lower dermal levels–allowed for a better capture of the data throughout the duration of the model. In addition, the parallel model shows a clear response to the experimental methodology of wiping lindane off the skin after the sixth hour.

Through the course of the PBPK modeling, we determined that although industry standards exist to predict absorption coefficients, some adaptations may need to be made depending on properties of the chemical and the vehicle of application. Furthermore, utilizing IVIVE techniques allows us to make predictions on experimental modifications that will allow for a stronger response to study design parameters, such as the duration of the experiment and frequency and timing of observations.