Title
Modeling and Simulation of Complex Biological Systems: Agent-Based Modeling of Hepatic Xenobiotic Elimination

Author
Sheikh-Bahaei, Shahab

Publication Date
2010

Peer reviewed|Thesis/dissertation

Modeling and Simulation of Complex Biological Systems: Agent-Based Modeling of Hepatic Xenobiotic Elimination

by

Shahab Sheikh-Bahaei

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Bioengineering

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

AND

UNIVERSITY OF CALIFORNIA, BERKELEY

Abstract

Hepatocytes interact with new compounds: metabolism is a frequent consequence. Hepatocyte micromechanisms have evolved to adjust in real time to interact with any never-before-seen compound. To demonstrate understanding of those mechanisms, next generation, pharmacologically useful hepatocyte models must strive to do the same: their micromechanisms recognize a compound as not previously seen and respond based on prior experience. How might that be accomplished when uncertainty is large and detailed data are chronically limited? My first goal was to answer that question: engineer a new class of hepatocyte models that draw on their own past experiences and compound physicochemical properties (molecular weight, pKa, logP, etc.) to respond uniquely to a new compound: the events that emerge during model execution provide a useful prediction of that compound's metabolic clearance. I used the new *synthetic* (combining elements to form a whole) method of modeling and simulation. The method involves building extant—actually existing and observable—working biomimetic micromechanisms. The micromechanisms within my successful *in silico hepatocytes* are comprised of autonomous, interactive objects and agents that map to six key components: extracellular media, cells, transporters, metabolic enzymes, cytosolic binding factors, and compounds. In silico clearance emerges from their interactions.

Within livers, hepatocytes are spatially organized into functional units called lobules. Those hepatocytes exhibit location dependent, possibly cooperative, properties, including gene expression and metabolic clearance. These properties frequently adapt to changes in compound type and exposure. Drug-induced hepatocyte damage can also be location dependent. Such spatially heterogeneous phenomena are referred to as hepatic zonation.

Next generation, pharmacologically useful, liver models must be capable of exhibiting similar phenomena under analogous conditions. My second goal was to extend my new class of models to the level of hepatic lobules. I organized autonomous agents, which map to small units of lobular function, into structures called Zonally Responsive Lobular Analogues (ZoRLA). The micromechanisms within each that enabled achieving a degree of validation were given two tasks: protect hypothetical external “tissues” by eliminating simulated toxins and minimize resource consumption. The zonation patterns that emerged are striking similarities to reported patterns. ZoRLA are designed to become components of future, virtual organisms.

To my parents, Eftekhar and Abdorrahim Sheikhbahaei,

my wife, Nasim, and my children, Sufie and Niksa.

Acknowledgement

This section has to begin with a heartfelt thank you to my wonderful wife Nasim Sassan. Her love and kindness has no boundaries. She has been with me much longer than graduate school. She declined her admission to doctor of pharmacy program at the University of New Mexico, so I can pursue my PhD at the University of California. She has made each day of this long journey special.

Of course, I owe a great deal to my advisor, Tony Hunt. I would like to thank him for his guidance and support. Tony's tireless dedication, clarity and precision of scientific thinking have been inspiring. I have learnt so much from Tony. I am grateful for the support that he has provided through rough spots in my graduate career.

I also want to thank my dissertation committee members, Donna Hudson and Leslie Benet, for their encouragement, support and valuable critique.

I thank Glenn Ropella for introducing me to the wonderful world of complex systems and computational biology. Meeting him in a workshop in Tuscan, Arizona led me to pursue my PhD in Bioengineering at the University of California, Berkeley & San Francisco.

Over the past five years the lab has been a second home to me. I have had a chance to work with a whole lot of great people in the lab. Thank you: Jesse, Jon, Li, Sean, Sunwoo, Teddy, and Laura. Pearl Johnson deserves a special mention. Although I had the pleasure of working with her for only a few months before her accident, I have never forgotten her kindness and thoughtfulness when I entered the lab. Her death was a great

loss to the lab.

I am indebted to my uncle, Mansoor Sheik-Bahae, for his guidance and advice. He helped me get situated in the United States when I moved in 2001.

I thank my sister, Shahrzad, for helping me with drug data collection, my brothers, Shahriar and Ali, for their support and encouragement. Shahriar has been a great source of biology and medical information for me too. I am grateful to my daughter, Sufie, for adding joy and happiness to my life during the years.

Also, I would like to thank my mother- and father-in-law, Zohreh and Abdolhossein Sassan, for their invaluable help and loving support. I would also like to thank my brothers- and sisters-in-law Maryam, Shadab, Ali and Shahab.

Lastly, I want to thank my parents, Eftekhar and Abdorrahim Sheikhbahaei, for getting me interested in science and for their encouragement. I am grateful for their enormous unconditional support throughout the whole process, and for taking pride in my accomplishments.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ANN: Artificial Neural Network
AUC: Area Under Concentration Curve
BCS: Biopharmaceutics Classification System
CA: Cellular Automata
CEC: Compound Elimination Count
CL: Clearance
CL_{int}: Intrinsic Clearance
CV: Central Vein
FCM: Fuzzy C-Means
FST: Fuzzy Set Theory
FURM: Functional Unit Representation Method
GA: Genetic Algorithm
GPCR: G-Protein-Coupled Recepto
HBA: Hydrogen Bound Acceptor
HBD: Hydrogen Bound Donor
IC: Ion Channel
ISH: In Silico Hepatocyte
LD: Lethal Dose
MR: Multiple Regression
MW: Molecular Weight
NR: Nuclear Receptor
PBPK: Physiologically Based Pharmacokinetic
PCPs: Physicochemical Properties
PEAF: Parameter Estimation via Analysis of Fuzzy Clusters
PK: Pharmacokinetic
PRN: Pseudo Random Number
PV: Portal Vein
RB: Rotatable Bond
SD: Standard Deviation
SM: Similarity Measure
SS: Sinusoidal Segment
TCDD: 2,3,7,8-tetrachlorodibenzo-p-dioixon
TPSA: Topological Polar Surface Area
ZoRLA: Zonally Responsive Lobular Analogues

1 Introduction

My research objective has been to develop and use new modeling and simulation methods to facilitate pursuit of deeper, more useful, mechanistic insight into processes involved in the hepatic clearance of xenobiotics in cultures and in the liver. I also sought more explanatory mechanistic models of observed, heterogeneous patterns of enzyme induction and localized toxicity caused by sustained exposure to some xenobiotics.

1.1 Scientific Modeling

The word “model” comes from the Latin word modellus (measure, standard). Generally speaking, a model is a simplified version of something real (e.g. a system or process). Some anthropologists believe that the ability to make conceptual models is the most significant quality, which gave Homo sapiens a competitive edge over concurrent hominid species (1). An example of early modeling work is Stone Age caveman paintings.

Science can be thought of as the quest for better models. Scientific models are born as hypotheses, some of which are falsified and eliminated by experimentation. The ones

that survive gradually evolve into theories. The process of scientific modeling is an iterative process involving four steps: 1) observations and measurements, 2) hypotheses (hypothetical explanations of observations and measurements), 3) predictions, and 4) experiments.

Different methods of reasoning have been identified as playing roles in scientific modeling (2): **Induction** is arrival at a conjecture (universal conclusion) based on a pattern observed in many particular cases; it is generalization: reasoning from detailed facts to general principles; it is a generalization drawn from patterns in observed data. **Deduction** is automatic and/or mechanical transformation of a set of statements; it is a purely mechanical (syntactic) transformation of the premises to a conclusion; it is transformation of assumptions into conclusions within a formal system [where if the assumptions are true, the conclusions and every intermediate step in the transformation are also guaranteed to be true]. **Abduction** is arrival at a conjecture based on a pattern observed in one or a few particular cases; it is construction of hypothetical speculations (consistent with current knowledge) about the process by which an outcome (phenomenon) came to be, where the hypotheses are all equally reasonable as long as they lead to the outcome; it is arrival at a conjecture (hypothesis) that would, if true, explain the relevant evidence.

1.2 Synthetic Modeling

A research objective is to develop better working hypotheses about the mechanisms that play roles in the emergence of biological phenomena when uncertainty is large and data are chronically limited. The synthetic method of modeling and simulation is a

relatively new experimental approach that can facilitate biomimetic mechanism discovery. Relying on object oriented programming and exercising abductive reasoning, the method involves building *extant*, working, software mechanisms that, upon execution, can exhibit biomimetic emergent phenomena. Building extant mechanisms is fundamentally different from the well-known approach of "modeling the data."

Using software engineering methods, we can create code that, when executed, produces mechanisms, which give rise to phenomena that are strikingly similar to specified phenomena. Yet there may be no logical mapping from event execution in the simulation to the biology. However, biologically inspired requirements can be imposed to shrink and constrain the space of software mechanism and implementation options that successfully exhibit those targeted phenomena. A continuation of that process can lead to extant software mechanisms (and phenomena) that are increasingly analogous to their biological counterparts. In so doing, we are not building a model based exclusively on known biological facts and assumptions, because the facts are often insufficient to do so. Furthermore, keeping track of all the assumptions and assessing their compatibility can become an unwieldy, time intensive task. Rather, we are exploring the space of reasonably realistic, biomimetic mechanisms that can cause the emergence of prespecified phenomena. The focus is on inventing, building, exploring, challenging, and revising plausible biomimetic mechanisms. To emphasize aspects of construction and method, specifically combining often varied and diverse elements, so as to form a coherent whole, I propose that the resulting models are synthetic *analogues*.

1.2.1 Design Guidelines

Determining Model Usage

The first consideration in any modeling effort is to determine why the models are being created, and identify situations in which they will be used (3). As explained in (2), synthetic modeling starts with specifying a list of *targeted attributes.* The modeling effort then strives to discover mechanisms that achieve those attributes. A requirement is that the models and their components be sufficiently flexible so that during future project extensions, or when other investigators use the models, they could be easily modified to account for an expanding list of attributes. Clear statements about use and targeted attributes facilitate selecting specifications. Clear specifications guide model design and development and help one avoid potentially unproductive tangents. Each attribute achieved provides a degree of validation.

Next, to achieve the targeted attributes, an iterative construction and refinement protocol is followed (2). The iterative refinement protocol cycles through the following eight steps.

1) Choose an initial, small subset of attributes to target.

2) Select a granularity level that will enable comparing measures of simulated and targeted attributes.

3) For each attribute targeted, specify a desired level of phenomenal similarity (e.g., within ± 25%). Approach in stages: begin with relaxed similarity measures.

4) Posit coarse-grained, discrete mechanisms that may generate analogous phenomena while requiring as few components as is reasonable.

5) Create logic to be used by each component. Instantiate components and mechanisms. Update specifications.

6) Conduct many, simulation experiments. Measure a variety of phenomena to establish *in silico*[1] to wet-lab similarity and lack thereof.

7) Tune (parameterize) to achieve analogue similarity specified at step 3. When the effort fails, return to step 4. When successful, return to step 3 and decrease the flexibility of the similarity measure.

8) Add one or more new attributes until the current analogue is falsified. Return to step 2. Strive to achieve the expanded attribute list with as little component reengineering as possible.

Building Extant Mechanisms

"What I cannot create, I do not understand." –Richard Feynman

Hepatic mechanisms are commonly described in the literature using diagrams combined with textual descriptions. These kinds of explanations are purely conceptual. Complex mechanisms are difficult to conceptualize. As a result, possible flaws in conceptual mechanisms may not be evident. Generally, conceptual mechanisms are difficult to falsify. Concrete mechanisms, on the other hand, are more easily challenged experimentally. I concluded that, in striving toward my objectives, it is essential to adopt as a strong guideline the dictum of the great physicist, Richard Feynman. When feasible, we must prefer extant (actually existing; concrete) over conceptual mechanisms.

[1] performed on computer or via computer simulation

Parsimony

"Everything must be made as simple as possible, but not simpler" – Albert Einstein

Parsimony is an additional strong guideline. My goal was to construct mechanisms that would be sufficiently complicated to achieve multiple, targeted phenotypic attributes, but no more so. In order to maintain parsimony, analogues should be constructed using components that are just fine-grained enough to produce targeted phenomena and accomplish the analogue's specified uses. My plan was to design analogues so that I could add additional detail easily when it was needed.

1.3 Synthetic Approach to Pharmacological Modeling

Pharmacology is the science of interactions between the body and exogenous chemicals (xenobiotics) that alter regular biochemical functions. Drug disposition and metabolism is an important aspect of pharmacology. Understanding the fate of drugs (pharmacokinetics) and their effects (pharmacodynamics) is challenging: 1) many mechanistic details remain to be discovered, 2) many of the details of posited mechanisms are poorly understood, 3) the considerable inter- and intra-individual variability is poorly understood and difficult to anticipate, and 4) pharmacological phenomena are a consequence of interventions by administered compounds on already existing and operating mechanisms.

Accordingly, to build useful synthetic models that can mimic pharmacological phenomena, the analogues will need to have the following characteristics. 1) Analogue components representing compounds should exist independent of mechanisms. The compound analogues need to carry information about the molecular structure they

represent. 2) In silico mechanisms should exist and function in the absence of a compound analogue intervention. 3) The model must be capable of exhibiting altered behaviors upon compound–analogue interaction. 4) Components of *in silico* mechanisms must be able to recognize compound analogues from their physicochemical properties. 5) In silico mechanisms must be able to respond to changes in dose and type of compound analogue.

The In Silico Liver (ISL) (4), is the first example of a synthetic, pharmacological model capable of exhibiting most of the above characteristics (5, 6). It was built to test and challenge concepts about liver function and the role of hepatic microenvironments in normal and disease states. An ISL maps to an *in situ* rat liver undergoing single pass perfusion. It represents a liver as an organized assembly of individually distinct primary units, either acini or lobules. It is sufficiently flexible to represent different aspects of hepatic biology at several levels of resolution (4).

To leverage the ISL, I focused on the properties of hepatocytes (the hepatic cells that most often respond to administered compounds), both *in vivo* and in cultures. The envisioned hepatocyte models were intended to evolve and become compatible with the ISL, so that they appropriately function if placed within the ISL, without having to be reengineered.

1.4 Focus on Metabolic Clearance By Hepatocytes

The mechanisms involved in development and maintenance of the spatially heterogeneous, homeostatic functions of liver are not fully understood (7). The liver is the primary organ responsible for drug and xenobiotic metabolism. Several traditional

liver models, including well-stirred compartments, the single tube, convection-dispersion, and interconnected tube models, have been developed for use in PK studies (8). However, because such models are data-centric, knowledge about hepatic structure and function are abstracted away. Therefore, the model components do not map well to biological mechanisms.

Hepatocytes are primarily responsible for clearing xenobiotics. The behaviors of hepatocytes in lobules are location dependent. Hepatocytes express heterogeneous, zone-specific enzymes and transporters, which results in portocentral gradients of gene expression. This heterogeneity, along with calcium-mediated, intercellular communication, suggests that hepatocytes may cooperate in some activities.

Hepatocytes individually choose strategies to clear compounds, i.e. they metabolize or ignore any new compound presented to them, apparently following an inherent agenda, the principles of which we do not fully understand. These assorted strategies might be based on their cumulative experience in responding to local environment and feedback provided by the larger whole of which they are part. If so, the collective behavior of the liver might be an *emergent* property of these distributed actions.

1.4.1 The In Silico Liver and Its Intended Uses

The In Silico Liver (ISL) (4, 9, 10) is a discrete, componentized, physiologically based, computational, liver analogue that is intended for refining, exploring, and testing hypotheses about interacting mechanisms that influence the transport, metabolism, and hepatic disposition of compounds of interest. An idea was that through the combined use of discrete and synthetic methods to more realistically model causal mechanisms in

the context of a realistic morphology, computational scientists will be able to better anticipate the PK consequence of changes in drug structure or pathology (4).

ISL execution is similar to conducting a wet-lab experiment. ISL uses therefore fall into two categories: 1) testing a hypothesis: execution of this ISL, configured and parameterized in this particular way, will produce phenomena that are measurably similar to the wet-lab observations of interest; and 2) exploring the consequences of some ISL perturbation. The expectation is that, if ISL-to-liver mappings are reasonable, then the mechanistic and phenomenal consequences of the ISL intervention will reasonably anticipate the outcome of such an intervention on a rat liver. Given these two categories, we can state specific ISL uses (2):

- Discover plausible mechanistic explanations 1) for drug disposition phenomena; 2) of hepatic responses to xenobiotic exposure; 3) of hepatic zonation phenomena, including clearance; 4) for differences and extents in drug disposition in normal and different types (and extents) of diseased and injured livers; 5) of disease progression; and 6) of causes of xenobiotic hepatotoxicity.
- Falsify or validate mechanistic hypotheses about clearance and/or disposition phenomena.
- Facilitate discovering plausible upward and downward linkages that will enable instantiating (represent by a concrete instance) details of genotype-phenotype linkage.
- Study plausible drug-drug interactions and predict their consequences.
- Explore and challenge 1) plausible mechanisms of cell-cell interactions and communication, and 2) plausible consequences of individual differences in hepatic

properties on phenomena of interest.

- Anticipate (predict) 1) aspects of the hepatic disposition of new compounds, and 2) plausible consequences of changes in gene expression.
- Provide validated subcomponents for experiments conducted in other labs studying other phenomena.
- Encourage and exercise abductive inference, which is essential in discovery and to the scientific method.

1.5 Specifications and Capabilities

To achieve the above and also to bridge "the gap" (4) between most current, system oriented, mathematical models and wet-lab models, the following general specifications are identified for the envisioned analogues:

- The analogues can be nested hierarchically. A component can be either atomic or composite. Atomic components, which have no internal structure, define the model's level of resolution—its granularity.
- An analogue must be comprised of quasi-autonomous, primary functional units (e.g., hepatic lobules). A functional unit can be comprised of other, quasi-autonomous functional units (e.g., a lobule consisting of cells).
- Enable straightforward inclusion of subcellular details, such as pathways and networks, when needed, without interfering with other micromechanisms or requiring system reengineering.
- Similar to that of the liver, a model's components should interact using only local information.

- The mappings from *in silico* models to the biology during experiments must be concretizable at three levels: 1) the mechanism components and their interactions, 2) how the consequences of multiple local micromechanisms merge at higher levels into measurable systemic behaviors, and 3) measurable similarity between multiple *in silico* attributes and corresponding wet-lab attributes.
- The inverse mapping from observable biological phenomena to generators (mechanisms) is one-to-many. The framework and methods must enable multiple, equally plausible, hepatic mechanisms, to operate in parallel with the objective of falsifying one or more as refinement progresses.
- The model's components must be linkable with other models and other model types.
- In silico mechanisms must be transparent. The micromechanisms of a falsified model need to be examined during simulations to see what, when, where, and why the failure occurred. So doing leads to new knowledge and new hypotheses.

In addition to the above specifications the following were specifically targeted for the models developed during my research.

Intelligent Analogues

In order to achieve the fifth requirement in Section 1.3, the analogues should be able to autonomously adapt to changes in the environment of which they are a part. They should be able to respond to new, previously unseen situations. For example one should be able to take a new compound, for which the model has not been validated, introduce it to a validated model, and observe the consequences on interactions, exactly as done with *in vitro* hepatocyte cultures. So doing presents several significant model-engineering challenges. The components in the validated model should contain mechanisms that

can recall how they interacted as they did with each of the already-encountered compounds. They need mechanisms that enable them to recognize and respond to the physicochemical properties of the new compound. Therefore, they need some primitive intelligence that enables them to use what they have learned and reparameterize themselves to respond to the new compound. Consequently, a goal of this work was to explore and provide options for that primitive intelligence.

Reusability in Different Contexts

The models are intended to be reused and iteratively improved. In other words, they should be flexible enough to function in different experimental contexts. It follows that given analogues (e.g., cells) that have achieved a degree of validation in one context (such as *in vitro*), it should be possible to move them to a new context (such as an ISL) and observe their behavior in that new context. When falsified, it should be straightforward for the models to relocate in mechanism space.

Taking validated *in vitro* hepatocyte analogues (developed in the first part of this work), and relocating them into a spatial arrangement that mimics a lobule cross section, faces an important context issue: hepatocytes *in vitro* all function alike, whereas in the liver they exhibit location dependent function, and the mechanisms responsible for that zonation behavior are still poorly understood. Consequently, the second part of this work focused on discovering one or more plausible mechanisms for dynamic hepatic zonation of drug clearance properties.

1.6 Thesis Outline

In Chapter 2, I present a stochastic agent-based method to instantiate, test, and

validate simulation models of cultured primary liver cells (hepatocytes). The models are referred to as *in silico hepatocytes* (ISH). The mechanisms involve interactions among objects and agents representing six key components: extracellular media, cells, transporters, metabolic enzymes, cytosolic binding factors, and drugs. Additional components can be easily added as needed. Each object acts autonomously and interacts with other objects according to a set of simple axioms, reflecting our knowledge of intracellular mechanisms. An *in silico* clearance property emerges from interactions of these objects and agents. The interactions take into account drug physicochemical properties (via molecular descriptors, such as molecular weight, pKa, logP, etc).

In Chapter 3, I use the synthetic method to build an agent-based model of liver zonation. Liver cells express heterogeneous, location-dependent enzyme and transporter activities to detoxify compounds, apparently following an intrinsic agenda, the principles of which are not fully understood. This phenomenon is known as liver zonation. In order to gain insight into this phenomenon, we developed and validated *in silico* agents that collectively are able to mimic hepatic location-dependent behavior. Using the synthetic method, we constructed extant, hepatomimetic analogues *in silico*: Zonally Responsive Lobular Analogues (ZoRLA). Portions of sinusoids containing one or more cells are modeled as autonomous agents arranged on a 2D grid. The arrangement mimics a lobule cross section through hepatic tissue. An agent's task is to protect hypothetical external "tissues" by eliminating simulated toxins. In addition, each agent strives to minimize its energy (internal resources) consumption while protecting the external tissue from being damaged. Each agent uses local information to choose and update a clearance strategy – the probability of eliminating an encountered simulated toxin. All

agents use the same simple optimization (learning) logic. The patterns that emerge have striking similarities with the observed zonation patterns within the liver. In addition, the results support the idea that creation and regulation of liver zonation requires at least two independent subsystems acting on two different types of signals.

2 In Silico Hepatocyte

We have used a stochastic agent directed, synthetic method to instantiate, test, and partially validate simulation models of cultured primary hepatocytes (the primary cell of the liver). In this chapter we focus on hepatocytes grown *in vitro* using a "sandwich" culture method that enables their properties and behaviors to more closely match those observed *in vivo* in intact laboratory rats. The models are referred to as *in silico* hepatocytes (ISH). The mechanisms involve interactions among objects that map to the key components: extracellular media, cells, intercellular tight junctions, intercellular lumen, transporters, metabolic enzymes, cytosolic binding factors, and drugs. The interactions take into account the physicochemical properties of the simulated drugs. The ISHs are designed for stand-alone experimentation; they can also function as components in hierarchical multi-models of larger systems such as a liver within a whole simulated organism. This chapter is organized as follows:

In Section 2.1 (adapted from (11) with minor revisions) we present the first version of the model, ISH1. We validated ISH1 using *in vitro* measures of cellular uptake and biliary clearance of four compounds (salicylate, taurocholate, enkephalin, and

methotrexate).

In Section 2.2 (adapted from (12) with minor revisions) we present a slightly improved ISH, called ISH2; and propose a method for estimating its parameter values. The method enabled ISH2 to interact with simulated drugs to reasonably match results from *in vitro* hepatocyte excretion studies. Further, we use the estimation method to reasonably anticipate the biliary transport and excretion properties of a new compound based on the acceptable parameter values for previously encountered compounds. We use Fuzzy c-Means (FCM) classification algorithm to determine the degree of similarity between previously tuned compounds and the new compound. Specifically, a set of simulation parameters for enkephalin was predicted using the tuned parameter values of salicylate, taurocholate, and methotrexate.

In Section 2.3 (adapted from (12, 13) with minor revisions) we generalize the method as a parameter estimation algorithm for agent-based models and apply it to predict *in vitro* human clearance of 73 drugs. The algorithm estimates model parameter values for new situations utilizing the characteristics of previously simulated conditions. We applied the algorithm to estimate parameter values of ISH2.

In Section 2.4 we added additional components and micromechanisms to the model and called it ISH3. We then estimated the parameters individually for each micromechanism by directly mapping the mechanisms to drugs physicochemical properties. The estimated parameter values were use to predict *in vitro* hepatic clearance of compounds.

2.1 In Silico Hepatocyte: Agent-Based Modeling of the Biliary Excretion of Drugs In Vitro[1]

2.1.1 Introduction

Modeling and simulation of biological systems is done both in continuous and discrete domains. Differential equations have been the tool of choice in the continuous domain. The behaviors and features of a biological system that are referred to collectively as its phenotype are too diverse and complex to be described by even very large sets of differential equations, and the apparent informal, stochastic nature of biological phenomena cannot be easily conveyed. In the discrete domain, *cellular automata* (CA) approaches have been used (12, 18) as tools for modeling complex collections of biological processes. The *lattice gas* models, also known as particle systems, comprise a well-known CA class. Usually driven by random events, these models consist of a discrete grid on which particles move about and interact with each other. When implemented in an object-oriented framework, the objects within the lattice can become independent software agents.

A class of biological models is presented in (6), which is based on the idea of "middle-out" constructive (synthetic) modeling strategy rather than the traditional top-down and bottom-up modeling and simulation approaches. Members of this class are referred to as *biomimetic in silico devices*. They are designed to generate behaviors that are useful analogues of a set of referent behaviors. The analogues are constructed from software components that are designed to map logically to biological components at

[1] Adapted from (11) with minor revisions

multiple levels of resolution, which facilitates modeling more complex biological phenomena.

Following the guidelines presented in (6), (4), and (19), and using an agent directed approach we propose a biomimetic device called *in silico hepatocyte* (ISH) to simulate attributes of hepatocytes (the parenchymal liver cells) grown in various *in vitro* environments.

A goal of this work is to develop an ISH (ISH1) that is sufficiently flexible to be used as a component in larger simulation models of more complicated *in vitro* and *in vivo* experimental systems such as the perfused rat liver discussed in (6) or the liver of a simulated patient. The design and structure of the current ISH is described in the Methods section. Its performance is demonstrated by using it to simulate the *in vitro* hepatic biliary excretion that can be observed and quantified using the specialized culture conditions described in (20): rat hepatocytes that have grown for 5 days in a "sandwich" culture are used to predict the *in vivo* biliary intrinsic clearance of drugs. The cumulative uptake of drugs by hepatocytes is measured under two different conditions: 1) standard, Ca-containing media and 2) media that is free of calcium ion (hereafter, Ca-free) for up to 10 minutes. The biliary excretion and clearance (CL_B) of each drug are estimated from the difference between the cumulative uptakes in the presence and absence of Ca^{2+}.

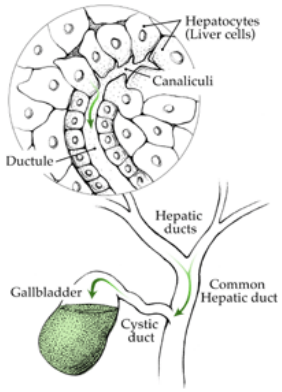

Figure 2.1 Hepatocytes excrete bile into canalicular spaces *in vivo* (21).

2.1.2 Biological Background

The liver can metabolize and excrete into bile many of the compounds and toxins that find their way into blood. This is an important step in their use by or elimination from the body. Bile passes into the small intestine and from there, a fraction of its content is reabsorbed and some of that may be ultimately eliminated by the kidney. The parenchymal cells of the liver, hepatocytes, excrete bile into intercellular spaces between themselves. These spaces merge to form bile canaliculi (Figure 2.1). In humans, the canaliculi merge and deliver their content to the gall bladder. In the *in vitro* sandwich-culture system, however, the bile can be sequestered in spaces (small lumens) that are created by adjacent hepatocytes that have formed tight junctions between themselves, as illustrated in Figure 2.2A. The tight junctions function as a seal between the luminal contents and the media external to the cells. The hepatocyte sandwich-culture system can be broadly subdivided into three spaces: intracellular (cytosol), canalicular lumen, and the incubation medium. In the system, Ca^{2+} is responsible for maintaining the barrier function of the tight junctions: they form a seal between the canalicular lumen and the incubation buffer. The barrier can be disrupted

by depletion of Ca^{2+} from the incubation medium. When such media is used, the solution (biliary secretions from hepatocytes) that had accumulated in the canalicular lumen mixes with the incubation medium. Therefore, the cumulative uptake when the standard media is used represents the amount of substrate both intracellular (cytosol) and in the canalicular lumen. However, when media that is Ca-free is used, the cumulative uptake represents the amount of substrate in cytosolic compartment only (20). Thus, the amount of substrate excreted in the canalicular lumen (i.e. *in vitro* counterpart to biliary excretion) can be estimated from the difference between the cumulative uptake in presence and absence of Ca^{2+}. The biliary excretion estimated by this method is consistent with *in vivo* biliary excretion (20).

2.1.3 Methods

We use agent-directed programming. We adopt the Functional Unit Representation Method (FURM) (4, 22) and framework, which makes use of three different models: an articulated, functional unit model (ArtModel); an accepted mathematical model—the reference model (RefModel); and an (*in vitro*) experimental data (DatModel) for validating the ArtModel. Within each simulation cycle these three models are executed by an *experiment agent* (ExperAgent). The ExperAgent is responsible for: managing the resources required for each experiment, controlling the models, taking data from the models, progressing from one experimental setup to the next, scoring each model against some performance measure, and acquiring telemetric data from the *in silico* experiments.

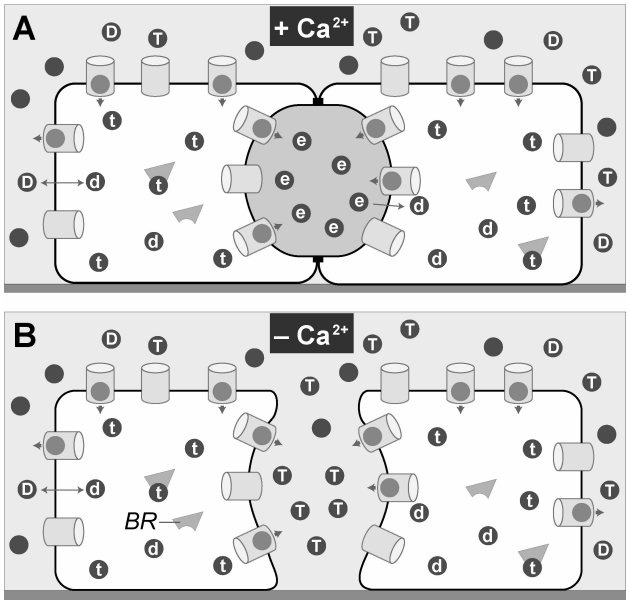

Figure 2.2 The sketch illustrates hepatocytes in a sandwich culture *in vitro* and the organization of the components within the ISH1. A: Two hepatic cells (white) are attached by tight junctions. A canalicular lumen space (shaded) is shown between them. The external medium includes Ca^{2+}. B: The system is the same as in A, but the Ca^{2+} in the media has been depleted breaking the tight junctions. All dark circles are SOLUTES. A blank circle (no letter) is an undesignated solute in the media space. Empty cylinders are transporter objects. A cylinder with light gray circle inside maps to solute being transported in the direction indicated by the arrow. br: binder object; e: solute that has been transported into the lumen space; t: SOLUTE that has been transported into a cell d: solute that has diffused into a cell; T: solute that has been transported out of a cell into the media; D: solute that has diffused out of a cell.

We represent hepatocytes using fixed agents placed in a 2D grid where mobile objects mapping to solute can interact with them stochastically (Figure 2.3). To avoid confusion hereafter and clearly distinguish *in vitro* components and features from corresponding *in silico* components and features, such as a "hepatocyte," a "solute," or "excreted," we use SMALL CAPS when referring to the *in silico* system. We model the canalicular lumen (center, Figure 2.2A) as an object acting as a container inside the simulated HEPATOCYTE into which SOLUTES can be EXCRETED by TRANSPORTERS. Simply eliminating this space simulates Ca^{2+} disruption of tight junctions and mixing of what would have been luminal contents with the extracellular media. A sketch identifying

several key features is shown in Figure 2.2.

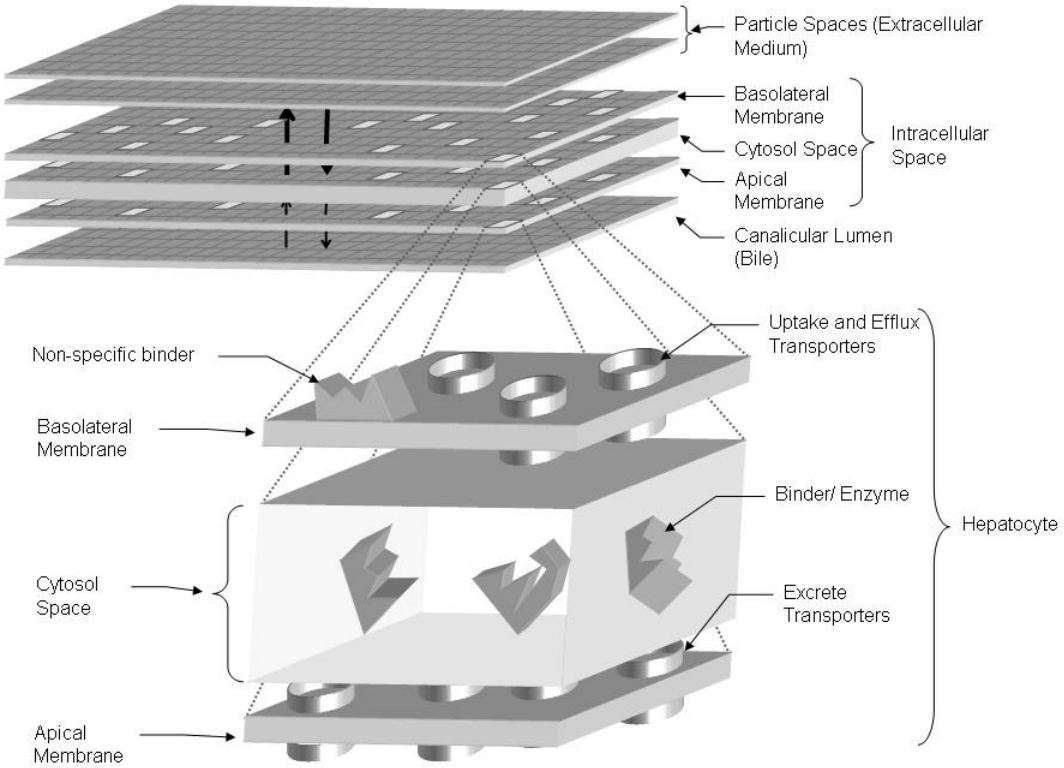

Figure 2.3. 3D illustration of the model. SOLUTES are initially placed in a Particle Space. They move stochastically and interact with HEPATOCYTES. Each hepatocyte may consist of BASOLATERAL and APICAL MEMBRANES, CYTOSOLIC space, TRANSPORTERS, non-specific BINDERS, and ENZYMES.

- *The Incubation Medium* is represented by a two dimensional square grid in which HEPATOCYTES and SOLUTES can be placed to interact with each other. SOLUTES move about using a Moore neighborhood.

- *Drug Compounds* (SOLUTES) are represented as independent, mobile objects that move around stochastically, governed by the flow of the incubation medium. During an experiment the event histories of SOLUTES (and other objects) can be tracked individually or as groups, such as SOLUTE that has been TRANSPORTED out of a CELL, or that has

diffused into a CELL.

- *HEPATOCYTES* are represented as shown in Figure 2.2. Each is constructed from objects that map to hepatocyte components and the environment: factors that can bind drug, enzymes, transporters, and a space for biliary excretion.

- A *Binder* is an object within HEPATOCYTES that can bind a free SOLUTE and hold onto it for a specified number of binding cycles.

- An *ENZYME* is a specialized form of binder. It can "metabolize" a bound SOLUTE by replacing it, following the binding period, with a metabolite object and destroying the replaced SOLUTE (for more details about these *in silico* Binders and Enzymes see (23, 24, 25).

- *TRANSPORTERS* belong to a subclass of binders. They can bind with free SOLUTE that is either inside or outside, and transport them to the opposite side of the CELL MEMBRANE, independent of the local SOLUTE concentration. When needed, TRANSPORTERS can be subdivided further into specialized forms. The following are three of the important TRANSPORTER parameters:

- *Transport-in/out_probability* specifies the probability that a TRANSPORTER will bind a given SOLUTE, once that SOLUTE is detected by the TRANSPORTER, and TRANSPORT it in or out of the CELL.

- *Binding_cycles* specifies the number of simulation cycles a SOLUTE will remain attached to a TRANSPORTER.

- *Excretion_space* stores excreted SOLUTES until they are removed to an

EXTRACELLULAR space. In this work, under standard culture conditions (with Ca^{2+}), they are not removed by outer spaces. They stay within this space simulating the fact they are "sealed" within a luminal space (designated e in Figure 2.2A). The average number of SOLUTES in this space is determined by the parameter *Excretion_Mean*, which is the mean of an exponential probability distribution. A SOLUTE in this space may diffuse back into the CELL (e-to-d in Figure 2.2A), depending on its physicochemical properties.

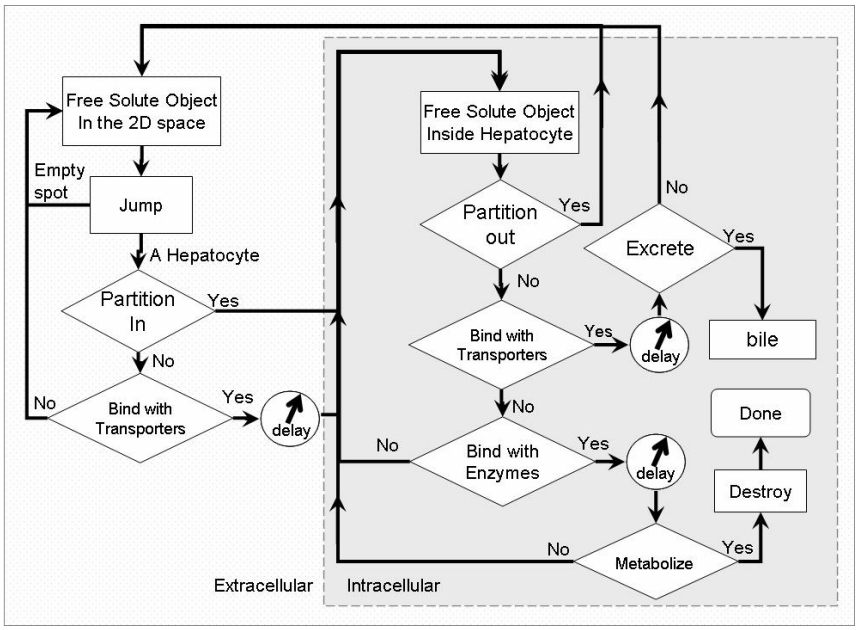

Figure 2.4. Trace of a SOLUTE (mapping to a drug particle) in the model.

In Silico Solute Kinetics

Figure 2.4 shows the trace of a SOLUTE in the simulation. SOLUTES are initially placed uniformly and randomly in the 2D space external to HEPATOCYTES. At each simulation cycle, a SOLUTE may stay in place or move randomly in one of eight directions (N, NE, E, SE, S, SW, W or NW with a probability of 1/9). A SOLUTE may, depending on its properties, partition into an encountered HEPATOCYTE. There is also a chance that

it may be transported (actively imported) into the CELL by TRANSPORTERS.

MEMBRANE Diffusion

Partitioning of solute into and out of hepatocyte cytosol is simulated as follows: when a free SOLUTE in the 2D space encounters a CELL, it may enter the CELL based on the values of two parameters:

Solute_Membrane_Cross-In_Probability and *Average_Cell_Capacity*. The former, which is governed by SOLUTE properties, is the probability that the SOLUTE enters the CELL passively. The latter is the mean of an exponential distribution that determines the number of objects a CELL can accommodate by passive TRANSPORT. The probability that a SOLUTE partitions into the CELL decreases as the number of intracellular SOLUTES increases. Each unbound intracellular solute may also partition out of the cell with a probability of *Solute_Membrane_Cross-Out_Probability*.

Active TRANSPORT

In silico, if a SOLUTE fails to enter the CELL by passive transport, it will be given a chance to bind with TRANSPORTERS that recognize it with probability of *transport-in_probability.* If recognized, it gets transported into the CELL. We assume that there are no spatially explicit arrangements of TRANSPORTERS within a CELL membrane.

In vivo, biliary excretion is performed by canalicular membrane transporters. In silico, as Figure 2.4 illustrates, once an intracellular SOLUTE binds to a TRANSPORTER, there is chance that the SOLUTE will get EXCRETED based on an exponential probability distribution with mean *Excrete_Mean*. If excreted, the object is put in a waiting list to

be removed by external spaces (such as a BILE CANALICULAR space when HEPATOCYTES are organized within a simulated hepatic lobule). If not removed, the list maps to solute "sealed" between HEPATOCYTES. If not excreted, the SOLUTE is placed in the transported-out list. Objects in that list are treated as if they had been transported out by the basolateral membrane TRANSPORTER, and are transferred to the 2D extracellular space (the simulated incubation medium) in the next simulation cycle.

In Silico Uptake and Efflux Studies

The cell culture media is represented by the 2D space; SOLUTE within HEPATOCYTES maps to substrate in the cytosolic compartment. SOLUTE in the *excretion space* maps to the excreted substrate in the canalicular lumen (bile). When the standard media was simulated, the *in silico* cumulative uptake was calculated using Eq. (2.1).

$$\text{Uptake}_{in\ silico} = \text{total of (partitioned-in + transported-in + excreted) SOLUTE} \quad (2.1)$$

When the Ca-free media is being simulated, the average number of excreted objects (*Excrete_Mean*) was set to zero to simulate the effect of Ca^{2+} depletion on the barrier function of tight junctions. Consequently, the *in silico* cumulative uptake for simulated Ca-free media could also be calculated by Eq**.** (2.1). The values of simulated efflux for both standard and Ca-free media were calculated using Eq. (2.2).

$$\text{Efflux}_{in\ silico} = \text{total of (partitioned-out + transported-out) SOLUTE} \quad (2.2)$$

Data Analysis

The *in vitro Biliary Excretion Index* is calculated using Eq. (2.3) (20). The same equation was used for the corresponding *in silico* calculation.

$$\text{Biliary Excretion Index} = (\text{Uptake}_{\text{standard}} - \text{Uptake}_{\text{Ca-free}})/\text{Uptake}_{\text{standard}} \quad (2.3)$$

Biliary clearance by the sandwich-cultured hepatocytes, $CL_{B(culture)}$ is calculated according to Eq. (2.4) (20).

$$\text{CL}_{\text{B(culture)}} = (\text{Uptake}_{\text{standard}} - \text{Uptake}_{\text{Ca-free}})/(\text{Time}_{\text{incubation}} \cdot \text{Concentration}_{\text{medium}}) \quad (2.4)$$

where $\text{Time}_{\text{incubation}}$ is the incubation time and $\text{Concentration}_{\text{medium}}$ is the initial substrate concentration in the incubation medium. The same equation was used to calculate the BILIARY CLEARANCE; the *in silico* SOLUTE density is defined as:

$$\text{In Silico density} = (\text{total solute})/(\text{total locations in the 2D space}) \quad (2.5)$$

Parameter Estimation

We used an optimization technique to estimate the parameters of this model. After each simulation experiment a similarity measure (SM) algorithm (26) assigns a similarity score to the output of the simulation. This score provides a measure of how similar the current output is to the referent experimental data. The goal is to maximize the SM. The optimization algorithm used the Nelder and Mead Simplex (27) method. The algorithm has been used widely (28, 29, 30, 31) for almost 40 years to solve parameter estimation problems. It is still the method of choice for many practitioners because it is straightforward to code and easy to use. The technique belongs to a class of methods that do not require derivatives and are often claimed to be robust for problems with discontinuities or where function values are noisy. This property makes it a good choice for helping to optimize our ISHs. There are several different versions and extensions. We are using the one described in (32) with minor changes. See Appendix A for more details on the algorithm.

The following parameters were used in the search: *Excretion_Mean*, *Average_Cell_Capacity*, *Solute_Transport-In_Probability*, *Solute_Transport-Out_Probability*, *Solute_Transport_Cycles*, *Solute_Membrane_Cross-In_Probability*, *Solute_Membrane_Cross-Out_Probability*, *Solute_Binding_Probability* and *Solute_Binding_Cycles.* For each of the four drugs a different set of parameter values was selected. Other parameters, listed in Table 2.1, such as *Space_Size*, *Hepatocyte_Density*, *Max/Min_Binders_per_Cell*, etc., were fixed for all four drugs. *Total_Solute_Particles* was calculated according to the *in vitro* concentration of the corresponding drug; see section 2.2.4 (below) for details. Table 2.2 shows the optimized parameter values for enkephalin and salicylate. In order to simulate the depletion of Ca^{2+} the *Excretion_Mean* was set to zero to essentially eliminate the excretion space.

2.1.4 Parameter Calculation Details

In vitro data of four drugs were obtained from (20). The incubation conditions are reported to be the same for all four drugs, however the concentration of the drugs varies from 1 µM to 15 µM. In order to be consistent with the *in vitro* experiments, the *in silico* relative ratio of DRUGS to HEPATOCYTES should be similar.

We define the following:

C_h: *in vitro* concentration of hepatocytes

C_s: *in vitro* concentration of solute

P: number of SOLUTES in the 2D space (or *Total-Solute-Particles*)

H: number of HEPATOCYTES in the 2D space

S: number of 2D spaces

d_h: density of HEPATOCYTES in the 2D space ($d_h = H/S$)

Obviously, the total number of HEPATOCYTES and SOLUTES should be less than the total number of grid spaces:

$$P + H \leq S \tag{2.6}$$

Also *in vitro*

$C_s = A_1/V$, $C_h = A_2/V \Rightarrow A_1/A_2 = C_s/C_h$

where C_s and C_h are the apparent concentrations of solute and hepatocytes respectively, A_1 and A_2 are the amounts of solute and hepatocyte, and V is the system volume.

$$P/H = \alpha \bullet C_s/C_h \tag{2.7}$$

where α is an *in vitro* to *in silico* unit conversion constant.

The problem is to choose α such that P and H each satisfies the above constraint for all four drugs.

Let $k_{max} = \max(C_s/C_h)$

$$P/H \leq \alpha k_{max} \tag{2.8}$$

On the other hand,

$P + H \leq S \Leftrightarrow P/H \leq (S/H) - 1 \Leftrightarrow P/H \leq (1/ d_h) - 1$

Consequently, the above constraint (Eq. (2.8)) will be satisfied if we choose α such that

$(1/ d_h) - 1 \leq \alpha k_{max}$

which implies

$\alpha \geq [(1/d_h) - 1]/max(C_s/C_h)$

In our case the C_h was 6.67×10^5 (cells/ml) for all 4 drugs, and the max C_s was 15 µM. Choosing $d_h = 0.2$ yields $\alpha \geq 1.78 \times 10^{11}$. Selecting $\alpha = 1.78 \times 10^{11}$(*cells*/*pmol*), P can be calculated by Eq. (2.7):

$P = 1.78 \times 10^{11} d_h S C_s(\mu M)/C_h(cells/ml)$.

2.1.5 Results

In Silico Biliary Excretion

The referent for this model is an *in vitro* system used for studying primary rat hepatocytes (20). Liu et al. (20) show that hepatocytes cultured in a collagen-sandwich configuration for up to five days establish intact canalicular networks, and reestablish polarized excretion of organic anions and bile acids. The system is a useful *in vitro* model for investigating the hepatobiliary disposition of compounds. The authors report that after the cells have been maintained in sandwich culture for five days, the cumulative uptake of [^{3}H] taurocholate (a common component of bile) by the hepatocytes was significantly higher in standard Ca-containing media, compared with that of Ca-free media. The difference is a consequence of accumulation of taurocholate in canalicular spaces. [^{3}H] Taurocholate efflux from cells pre-loaded with drug for five days was greater in Ca-free compared with standard Ca-containing media.

Drug Uptake

Figure 2.5 shows the uptake of four drugs in well-established, sandwich-cultured hepatocytes using both standard and Ca-free media. Also shown are the *in silico*

UPTAKE values using ISH1 optimized for each of the four drugs. The values were calculated using equations (2.1), (2.3) and (2.4).

Table 2.1 Calculated and fixed parameter values

Parameter		**Enkephalin**	**Salicylate**
In vitro	Cell density (cells/ml)	6.67×10^{5}	6.67×10^{5}
	Drug concentration (M)	1.50×10^{-5}	1.00×10^{-6}
α (unit conversion constant)		1.78×10^{11}	1.78×10^{11}
In silico	*Space Size*	53 x 54	53 x 54
	Hepatocyte Dencity	0.2	0.2
	Min_Binders_per_Cell	5	5
	Max_Binders_per_Cell	10	10
	Min_Transporters_per_Cell	5	5
	Max_Transporters_per_Cell	10	10

Table 2.2 Optimized parameter values found for enkephalin and salicylate in standard buffer

Parameter	**Enkephalin**	**Salicylate**
Total_Solute_Particles	2290	153
artHepExcretionMean	0.008	0.0046
artCellAverageCapacity	0.16	0.01
artSoluteTransportInProb	0.0046	1 x10-6
artSoluteTransportOutProb	0.040	1 x10-6
artSoluteTransportCycles	1	2
artSoluteMembraneCrossInProb	0.0078	0.058
artSoluteMembraneCrossOutProb	0.144	0.207
artMetabolizationProb	0	0
artSoluteBindingProb	0.002	0.067
artSoluteBindingCycles	3	3

The duration of the *in silico* experiments was 20 cycles. Each unit of simulation time was 2 cycles. Figure 2.6 shows the correlation of *in silico* and *in vitro* Biliary Excretion Index and Biliary Clearance of compounds. We suggest that the *in vitro* and *in silico* values in Figure 2.5 are experimentally indistinguishable because the *in silico* values are within the range of variability that is seen for repeated wetlab experiments.

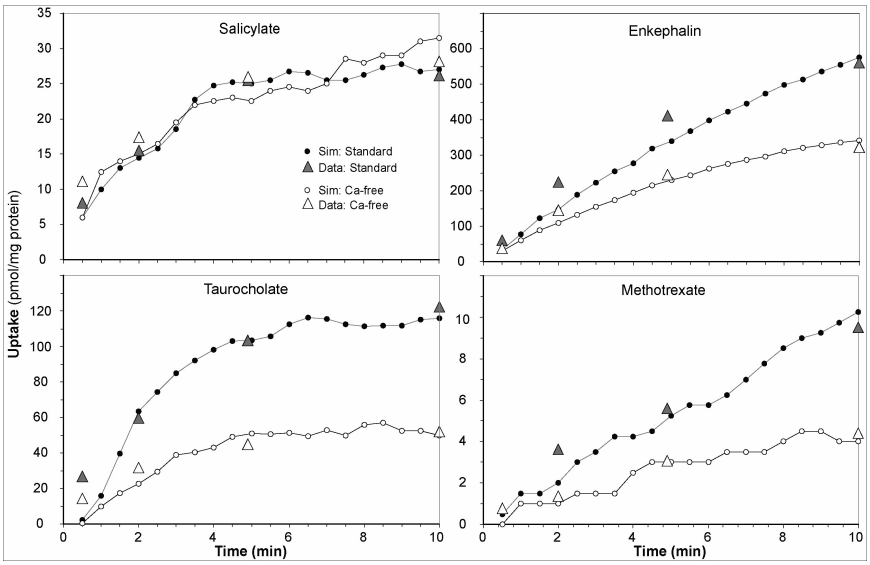

Figure 2.5 Simulated and *in vitro* outputs are shown for four compounds studied in standard and Ca-free media. The legend within the Salicylate plot frame also applies to the other three plots.

2.1.6 Conclusion

Using an agent based, constructive approach, we have presented and partially validated a set of simulation models for uptake and biliary secretion of compounds by hepatocytes grown *in vitro*. We have demonstrated how this model can be used to simulate the *in vitro* biliary excretion of drugs by hepatocytes grown in a sandwich culture.

The models are instantiations of the mechanism hypothesized by Liu et al. (20). Consequently, our *in silico* experimental results provide direct evidence that, at the low level of resolution used, the mechanism is an accurate representation of the actual *in vitro* events.

Although the parameters do not map directly to measurable biological counterparts, they can be estimated for a new drug using machine-learning tools such as Fuzzy Logic, Neural Networks etc. One of the important, future tasks is to demonstrate how this

model can be used to predict the biliary clearance of drugs.

The goal of scientific, biomedical simulation, in contrast to engineering simulation, is to discover plausible mechanisms for how a system might generate specific behavior. In cases where many of the elements of a process are unknown or unclear, we can build families of simulations that circumscribe an *in silico* behavior space that partially overlaps the behaviors of the referent system. When building such simulation families, many of the parameter values can be taken from or enlightened by data from biological referents. However, many parameters remain artificial or abstract. In the latter case they provide flexibility and allow more control over the search of the model behavior space. When the behaviors of these models and the referent biological system begin to converge, analysis of the artificial parameters is expected to help researchers generate new hypotheses for those parts of the system that are not explicitly available for study in the biological experiments.

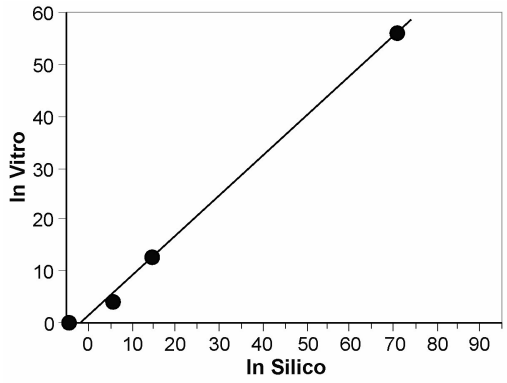

Figure 2.6 Correlation of *in silico* and *in vitro* biliary clearance. Circles show the calculated values from simulation (*in silico*) results vs. *in vitro* ($R^2 = 0.997$).

2.2 Prediction of In Vitro Hepatic Biliary Excretion Using Stochastic Agent-Based Modeling and Fuzzy Clustering[1]

2.2.1 Introduction

Accurate estimation of a drug's hepatic disposition (including hepatic metabolism, protein binding, intracellular sequestration, and biliary excretion) is a crucial step in the development of clinically practical drugs. Early in the new drug development process there is often a considerable over supply of candidates. Suitable *in silico* methods can help narrow the list of candidate compounds to a manageable number before beginning expensive wet-lab evaluation and screening of those selected (33, 34). Compounds that are likely to have undesirable metabolic and excretory properties need to be culled from the list. The current most widely used *in silico* methods rely on correlational techniques rather than estimates based on knowledge of the mechanisms involved. We focus on modeling the biliary excretion of compounds at the mechanistic level. Biliary excretion is a relatively complex process involving translocation across the sinusoidal membrane, movement through the cytoplasm, and transport across the canalicular membrane. Different transporters can be involved, as can metabolism. Intracellular spatial organization may be important. Competitive interactions can occur between these components and other compounds undergoing the same processes.

Numerous *in vitro* systems (e.g., isolated perfused livers, isolated hepatocyte, short-term cultured hepatocyte couplets and long term sandwich cultured hepatocytes) have been used to investigate biliary excretion. In the *in vitro* sandwich-culture system,

[1] Adapted from (10)with minor revisions

bile can be sequestered in spaces (small lumens) created by adjacent hepatocytes that have formed tight junctions between themselves, as illustrated in Figure 2.2A. The tight junctions function as a seal between the luminal contents and external media. For the low resolution model described herein, the system has been broadly subdivided into three spaces: intracellular (cytosol), canalicular lumen, and the incubation medium. In the *in vitro* system, Ca^{2+} is responsible for maintaining the barrier function of the tight junctions: they form a seal between the canalicular lumen and the incubation buffer. The barrier can be disrupted by depletion of Ca^{2+}. When that is done, the solution (biliary secretions from hepatocytes) that had accumulated in the canalicular lumen spaces mixes with the incubation medium. The cumulative uptake when the standard media is used represents the amount of substrate both intracellular (cytosol) and in the canalicular lumen. However, when Ca-free media is used, the cumulative uptake represents the amount of substrate in cytosolic compartment only (20). Thus, the amount of substrate excreted in the canalicular lumen (i.e. *in vitro* counterpart to biliary excretion) can be estimated from the difference between the cumulative uptake in the presence and absence of Ca^{2+}. We constructed a low-resolution (few components; limited detail) model to test hypotheses about the mechanistic details of biliary transport *in vitro* and to predict transport and excretion properties of newly encountered compounds. The biliary excretion estimated by this method for several compounds is consistent with *in vivo* biliary excretion data (20).

2.2.2 The Model

We use agent-directed programming to construct the model. Within the *in silico*

hepatocytes (ISH) we represent cells using fixed agents placed in a 2D grid where mobile objects mapping to solute can interact with them stochastically. The ISH used here (ISH2) is the very similar to that presented and described previously in Section 2.1. The descriptions that follow are consequently brief, yet sufficient for clarity. Consult Section 2.1 for additional detail.

To avoid confusion hereafter and clearly distinguish *in vitro* components and features from their *in silico* counterparts, such as a "hepatocyte," a "solute," or "excreted," we use SMALL CAPS when referring to the *in silico* system. We model the canalicular lumen (center, Figure 2.2A) as an object acting as a container inside the simulated HEPATOCYTE into which SOLUTES can be EXCRETED by TRANSPORTERS. Simply eliminating this space simulates Ca^{2+} disruption of tight junctions and mixing of what would have been luminal contents with the extracellular media, as illustrated in Figure 2.2. *The Incubation Medium* is represented by a 2D square grid in which HEPATOCYTES and SOLUTES can be placed to interact with each other. Drug Compounds (SOLUTES) are represented as independent, mobile objects that move stochastically (using a Moore neighborhood), governed by the flow of the incubation medium. During an experiment, the event histories of SOLUTES (and other objects), such as SOLUTE that has been TRANSPORTED out of a CELL, or that has diffused into a CELL, can be tracked individually or as groups. SOLUTES are initially placed uniformly and randomly in the space external to HEPATOCYTES. At each simulation cycle, a SOLUTE may stay in place or move randomly in one of eight directions (N, NE, E, SE, S, SW, W or NW with a probability of 1/9 each). A SOLUTE may, depending on its properties, PARTITION into an encountered HEPATOCYTE. There is also a chance that it may be transported (actively imported) into the CELL by

TRANSPORTERS. The pseudorandom numbers were generated from a uniform distribution using a Mersenne Twister random number generator.

HEPATOCYTES are constructed from objects that map to corresponding hepatocyte components and the environment: factors that can bind DRUG, ENZYMES, TRANSPORTERS, and a space for BILIARY excretion. A BINDER is an object that can bind or sequester free SOLUTE and hold onto it for a specified number of binding cycles. An *ENZYME* is a specialized form of BINDER. It can METABOLIZE a bound SOLUTE by replacing it, following the binding period, with a METABOLITE and then destroying the replaced SOLUTE (for more details see Section 2.1, and (6, 25)). *TRANSPORTERS* belong to a subclass of BINDERS. They can bind a free SOLUTE that is either inside or outside a CELL, and transport it to the opposite side of the CELL MEMBRANE, independent of the local SOLUTE density. Three important TRANSPORTER parameters are *Transport-in/out_probability* (it specifies the probability of binding a given SOLUTE), *Binding_cycles* (specifies how many cycles a SOLUTE remains attached), and EXCRETION_SPACE (the location of excreted SOLUTES until removal to an EXTRACELLULAR space). In cultures with Ca^{2+}, SOLUTES in the EXCRETION_SPACE are not removed simulating that they are "sealed" within a luminal space (e in Figure 2.2A). *Excretion_Mean* determines the average number of SOLUTE in this space. A SOLUTE within the space may move back into the CELL (e-to-d in Figure 2.2A), depending on its physicochemical properties.

In Silico Dynamics

Two parameters, *Solute_Membrane_Cross-In_Probability* and *Average_Cell_Capacity*, determine when a free SOLUTE in the "INCUBATION MEDIUM" space that has

encountered a cell may enter it. The former, which is governed by SOLUTE properties, is the probability that the SOLUTE enters the CELL passively. The latter, the mean of an exponential distribution, determines the number of objects a CELL can accommodate by passive transport. The probability of partitioning into a CELL decreases with increasing INTRACELLULAR SOLUTE. Each unbound INTRACELLULAR SOLUTE may also partition out with a probability of *Solute_Membrane_Cross-Out_Probability.*

If a SOLUTE fails to enter the CELL by passive transport, it will be given a chance to bind, with probability of *transport-in_probability*, with a transporter by which it is recognized. If recognized, it is transported into the CELL. We assume that TRANSPORTERS are placed randomly within a CELL MEMBRANE.

Once an INTRACELLULAR SOLUTE is bound, there is chance that it will get excreted based on an exponential probability distribution with mean *Excrete_Mean*. If excreted, it will be removed by external spaces. If not removed, the list maps to solute "sealed" between hepatocytes. If not excreted, the SOLUTE is treated as if it had been TRANSPORTED out by the basolateral MEMBRANE TRANSPORTER, and is transferred back to the 2D extracellular space (the simulated incubation medium) in the next simulation cycle. SOLUTE within HEPATOCYTES maps to substrate in the cytosol. SOLUTE in the EXCRETION SPACE maps to material excreted into the canalicular lumen (bile). For both standard and Ca-free media, the *in silico* cumulative UPTAKE was calculated by Eq. (2.1).

When the Ca-free media is being simulated, the average number of excreted objects (*Excrete_Mean*) was set to zero to simulate the effect of Ca^{2+} depletion.

Parameter Tuning

The ISH2 parameter space consists of several different parameter types, some of which directly map to measurable biological counterparts. However, most of the parameters are simulation-specific with no direct biological significance. The ISH2 behavior space of the model partially overlaps with the behaviors of the referent system when parameters have been appropriately tuned (adjusted). By so doing, we make the ISH2 behavior space converge on the behavior space of the referent system. In general, the parameter hyperspace of this model consists of both biologically meaningful (e.g. cell/compound density) and simulation-specific parameters (e.g. binding or membrane-crossing probabilities) that enable simulation of targeted behaviors. Whereas the simulation-specific parameters have no particular constraints, the biologically meaningful parameters are constrained to stay within reasonable values (so that, at a minimum, outcomes are consistent with cell life). Optimizing the whole parameter vector makes it easier to find solutions within the constrained regions. A failure of optimization might be caused by false (biologically unrealistic) mechanisms implemented in the model. On the other hand, a successful optimization provides some measure of validation for the implemented mechanisms. Doing so, however, is not a goal of this paper: we have already validated this model against data for four compounds (Section 2.1).

This parameter tuning can be done by optimization methods. Since this model is stochastic and therefore has discontinuity and noise in its behavior space, ordinary gradient-based Newton and quasi-Newton optimization techniques, which are commonly used for differential equations parameter tuning, cannot be applied. Non-gradient-based

methods such as GAs (genetic algorithms) or Nelder and Mead Simplex method (27) would be more successful.

After each simulation experiment a similarity measure (SM) algorithm (22) assigns a similarity score to the output of the simulation. This score provides a measure of similarity of the current output to the referent experimental data. The goal is to maximize the SM score. The optimization algorithm used is the Nelder and Mead Simplex (27) method that has been widely used (28, 29, 30, 31) to solve optimization problems with discontinuities or where function values are noisy. Among the several different versions and extensions of this method, we used the one described in (32) with minor changes (see Appendix A for details).

Parameter Estimation

In this section we present an algorithm which uses FCM to estimate the simulation parameters of enkephalin knowing the tuned parameters of salicylate, taurocholate, and methotrexate. Three compounds are minimal. It is, however, sufficient to demonstrate the approach. In an industry setting, data on many more compounds will be available. The four compounds were classified to two and three clusters using Fuzzy c-Means algorithm based on a selection of physicochemical properties that are expected to contribute to biliary clearance of the compounds. Initially the following physicochemical properties were considered: molecular weight, logP, hydrogen bond donor count, hydrogen bond acceptor count, rotatable bond count, tautomer count, pKa, TPSA, volume, GPCR ligand, ion channel modulator, kinase inhibitor, and nuclear receptor ligand. The results are shown in Table 2.3. See Appendix B for details on Fuzzy c-Means algorithm.

Table 2.3 Fuzzy classification results of salicylate, taurocholate, enkephalin and methotrexate based on their physicochemical properties (Table 2.3). C: number of clusters.

C	Group	Sal.	Taur.	Meth.	Enkeph.
2	1	0.9862	0.0736	0.2308	0.1529
	2	0.0138	0.9264	0.7692	0.8471
3	1	0.9981	0.0997	0.0492	0.0096
	2	0.0011	0.5492	0.8639	0.0291
	3	0.0007	0.3511	0.0869	0.9614

Table 2.4 Physicochemical properties of salicylate, taurocholate, methotrexate and enkephalin.

Property1	Sal.	Taur.	Meth.	Enkeph.
MW	140.1	515.7	454.4	645.8
logP	2.24	0.01	-1.28	2.01
HBD count	2	5	5	7
HBA count	3	7	12	8
RB count	1	7	9	7
Tautomer count	4	2	24	32
pKa	2.97	1.8	4.7	10
TPSA	57.5	144.1	210.6	199.9
Volume	119.1	483.1	387.4	569.7
GPCR ligand	-0.44	-0.26	0.22	-0.19
IC modulator	-0.08	-0.15	0.02	-1.05
Kinase inhibitor	-0.65	-0.47	0.11	-0.84
NR ligand	-0.58	-0.08	-0.36	-0.58

The results show that when divided into two groups, taurocholate, enkephalin, and methotrexate have more membership in the same group while salicylate belongs to another. However, when divided to 3 groups, taurocholate and methotrexate have membership in the same group while enkephalin and salicylate belong to different

[1] MW: molecular weight; HBD: hydrogen bond donor; HBA: hydrogen bond acceptor, RB: rotatable bond, TPSA: topological polar surface area, GPCR: G-protein-coupled receptor, NR: nuclear receptor, IC: ion channel. Property values were obtained from the following sources. http://www.molinspiration.com/cgi-bin/properties; http://www.syrres.com/esc/est_kowdemo.htm; and http://ibmlc2.chem.uga.edu/sparc/index.cfm.

groups.

Knowing the similarity of the compounds in the physicochemical domain we approximated their model parameter values assuming that similarity in the physicochemical domain maps to similarity in the model parameter domain. A set of simulation parameters is acceptable for a compound if it produces simulated results that are *good enough* for the questions or research task at hand, i.e. facilitate decision making during early drug selection and development. The Similarity Measure algorithm is used to measure the quality of the output. Given the preceding assumption, the parameter vector of compound *X* is estimated as the weighted average of the *in silico* parameter vectors of the other compounds. The proportional relevance of the parameter vector of compound Y, depends on the degree of similarity of *X* and *Y*.

In general, for a data set, S, containing n compounds S= $\{c_1, c_2,\ldots, c_n\}$, the PEAF algorithm (page 51) was used to estimate the simulation parameters of a new compound, c_{n+1}.

The accuracy of this estimation depends, of course, on how many compounds similar to compound X exist in the data set.

2.2.3 Results

The referent for this model is an *in vitro* system used for studying primary rat hepatocytes (20). Liu et al. (20) show that hepatocytes cultured in a collagen-sandwich configuration for up to five days establish intact canalicular networks and reestablish polarized excretion of organic anions and bile acids. The system is a useful *in vitro* model for investigating the hepatobiliary disposition of compounds. The authors report

that after the cells have been maintained sandwich culture for five days, the cumulative uptake of [^{3}H] taurocholate (a common component of bile) by the hepatocytes was significantly higher in standard Ca-containing media, compared with that of Ca-free media. The difference is a consequence of accumulation of taurocholate in intracellular and canalicular spaces established during the prior culture period. [^{3}H] Taurocholate efflux from cells pre-loaded with drug for five days was greater in Ca-free compared with standard Ca-containing media. There is, of course, variability both within and between experiments.

Figure 2.7 shows the correlations between *in vitro hepatocyte* and *in silico* ISH2 uptake values at different times. The *in vitro* values were obtained 20, 24) using well-established, sandwich-cultured hepatocytes using both standard and Ca-free media. The *in silico* UPTAKE values were calculated using the equation Eq**.** (2.1). The ISH2 parameter values for each drug were iteratively optimized as reported in Section 2.1.

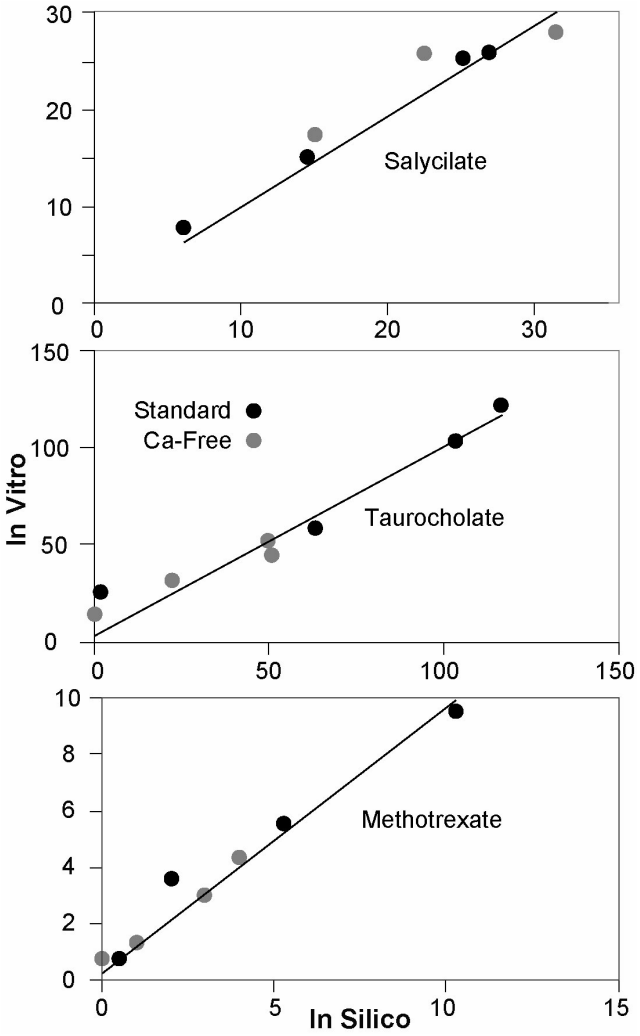

Figure 2.7 Correlation between *in vitro* and *in silico* uptake values at four different times for standard (black circles) and Ca-free media (gray circles). In vitro the times were 0.5, 2, 5 and 10 minutes.

The measured uptake of enkephalin in well-established, sandwich-cultured hepatocytes, using both standard and Ca-free media, is presented in Figure 2.9A. Also shown is the model-predicted time course of *in vitro* enkephalin UPTAKE under those same two conditions. In Figure 2.8B the matches to both types of uptake are shown when model parameter values are iteratively tuned Section 2.1. Note that the *in vitro* uptake of enkephalin is about four times that of taurocholate, which has the largest uptake of the three compounds). The predicted uptake of enkephalin in the Ca-free buffer

reaches its steady state too early. This might be due to the saturation of the uptake transporters used in the model. The *in silico* concentration of enkephalin is much higher than for the other three, and that may be a reason for saturation.

Table 2.5 Tuned vs. Estimated parameter values for Enkephalin.

Parameter	Tuned	Estimated
ExcretionMean	1.20	0.51
BindersPerCellMin	5	5
BindersPerCellMax	10	10
CellAverageCapacity	0.19	1.15
TransportersPerCellMin	5	5
TransportersPerCellMax	10	10
SoluteTransportInProb	0.016	0.023
SoluteTransportOutProb	0.0808	0.14
SoluteTransportCycles	1	2
SoluteMembraneCrossInProb	0.012	0.035
SoluteMembraneCrossOutProb	0.095	0.37
MetabolizationProb	0	0
SoluteBindingProb	0.032	0.052
SoluteBindingCycles	4	3

2.2.4 Conclusion and Discussion

In Section 2.1 we present an earlier version of the ISH2 that is, upon properly tuning its parameters, capable of mimicking its biological referents. In this section, we use ISH2 to *predict* the behavior of the referent system when introduced to a new compound not previously encountered. In order to predict that behavior, we needed means to *estimate* an appropriate set of parameter values. In this work we used an algorithm (the PEAF algorithm) to estimate those values. It utilizes Fuzzy c-Means to cluster previously encountered compounds based on their physicochemical properties (Figure 2.8). Our FCM approach offers three important advantages:

1. Because FCM is an unsupervised learning method (it does not need to be trained

with training data to work) the problem of over-fitting to the training data is minimized. That is particularly important in our case due to the small size of our data set.

2. It provides soft clusters. They are more biologically realistic than the hard crisp clusters.

3. Unlike other classification methods (such as ANNs) the FCM only needs to specify a few parameters. In fact, in the algorithm presented, it has only *m* and *G*, where we chose *G* = I (the Identity matrix).

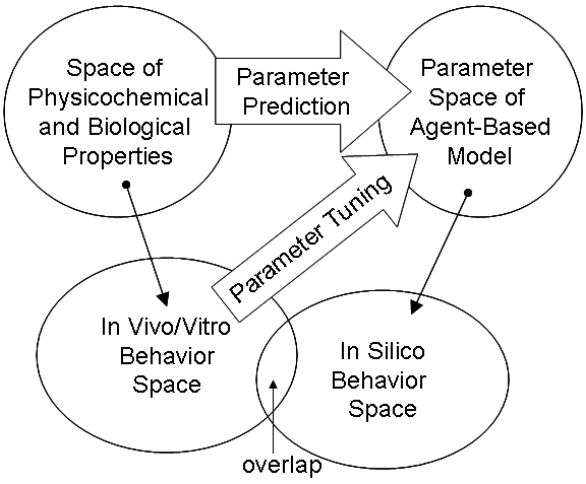

Figure 2.8 The ISH2 parameter space consists of both biologically meaningful and simulation-specific parameters. The ISH2 behavior space partially overlaps that of the referent system. By properly adjusting parameters, we make the ISH2 behavior space converge on the behavior space of the referent system. Parameter prediction is a direct mapping from physicochemical properties to the ISH2 parameter space. Parameter tuning draws its information from the biological behavior space.

We now discuss acceptability of the results in Figure 2.9: Within experiments, there can be orders of magnitude differences between active and passive uptake of different compounds along with comparable differences in biliary excretion. Typically, the uptake and biliary excretion values of the same compound, between experiments, using

essentially the same experimental system, is within a factor of two. The results of simulations can be more precise, but without drawing on additional information, their predictive accuracy can not be better. The simulations described here are intended to be analogous to repeat *in vitro* experiments: an observer should not be able to distinguish between data coming from a repeat *in vitro* experiment and data from an experiment on a tuned ISH2. With those model use expectations set, the results in Figure 2.9 are minimally acceptable. They are *good enough* to facilitate decision making during early drug selection and development. The simulated enkephalin results are based on the uptake data of only three other compounds. Our expectation is that when the above procedure is used with descendants of model in Figure 2.2 to predict the uptake and excretion properties of future new compounds, the usefulness of predicted results will improve with each expansion of the set of successfully represented compounds.

The traditional approach to predicting the *in vitro* and *in vivo* properties of new compounds is to search for patterns within large data sets of measured biological property data and then to seek patterns within the set of compound property values of those compounds for which correlations exist. Knowledge about the mechanisms that generated the biological data is only used indirectly. A contribution of this section is offering a method for combining both the knowledge of mechanisms and the patterns found in the space of the physicochemical and biological properties. The models and approach described here are designed to leverage that knowledge by representing and improving our understanding of the generative relationships within the target biological system. The generative relationships between components within the ISH2 stand as a hypothesis of how the corresponding *in vitro* phenomena may be generated. As such,

the hybrid approach presented here is expected to significantly improve our ability to anticipate the biological properties of compounds of interest. The approach is new: more work is needed to uncover and understand limitations and to delineate advantages relative to other methods typically used (optimization, data fitting etc.).

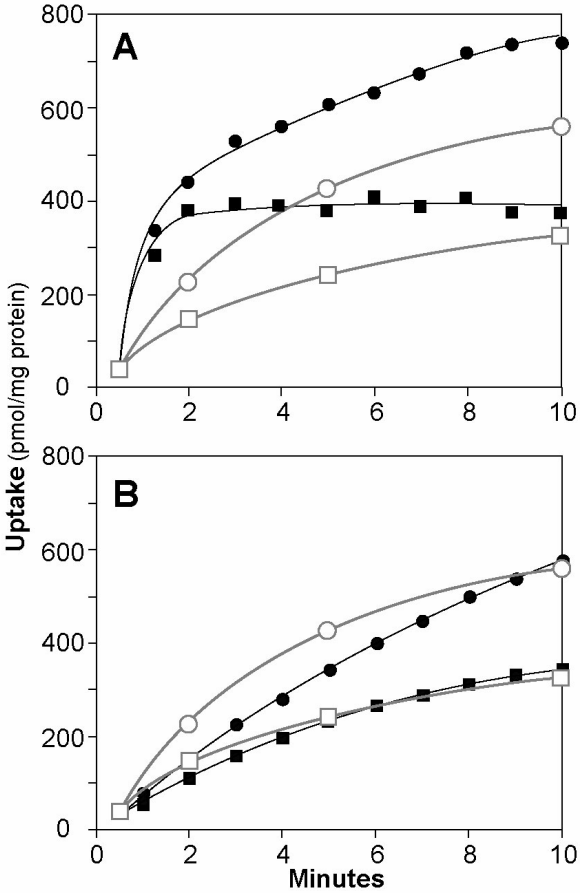

Figure 2.9 Comparisons of enkephalin uptake, *in vitro* and simulated, for two different growth conditions; the simulations use either the estimated (A) or tuned (B) parameter values (Table 2.5). Within each graph the simulated results (black symbols) are contrasted to the *in vitro* data (unfilled, gray symbols). Circles are cumulative uptake and biliary excretion values in standard media (with Ca^{2+}) *in vitro* (gray) and *in silico* (black). Squares are cumulative uptake values in Ca-free media *in vitro* (gray) and *in silico* (black). The simulated results were generated using the iteratively tuned ISH2 parameter set reported in Section 2.1.

2.3 Parameter Estimation via Analysis of Fuzzy Clusters (PEAF): An Algorithm to Estimate Parameters of Agent-Based Models[1]

2.3.1 Introduction

Agent-based modeling is being used in a variety of fields: examples include social sciences (35, 36, 37, 38, 39), supply chain optimization and logistics; modeling of consumer behavior; distributed computing; workforce management; traffic management; portfolio management; complex systems, artificial life, genetic programming and genetic evolution 40, 41, 42, 43, 44, 45); bacterial chemotaxis signaling pathways (45, 46); population ecology (47, 48, 49); social and economic systems (43, 50, 51); and cellular behavior (24, 25, 52, 53, 54).

Agent-based models commonly require many parameters. Together, they determine the global dynamics of the system. Small changes made to one parameter can lead to an important change of the dynamics of the entire system. Consequently, identifying informative and plausibly realistic regions of parameter space for exploration can be time-intensive (55). Several automated techniques have been used, including the Nelder and Mead Simplex Method (27) and Genetic Algorithms (55). Once parameter vectors have been identified that are suitable for several situations, one can become interested in predicting system behavior for a new situation. In this paper we propose a method to estimate such parameters based on previously seen cases in order to predict system

[1] Adapted from (1213) with minor revisions

behavior for a new situation. The proposed method uses the Fuzzy-c-Means (56) classification algorithm. As a proof of concept we apply the method to an agent-based model of hepatocytes, and make predictions.

2.3.2 Methods

Parameters of Agent-Based Models

Parameters in biologically focused, agent-based simulation models can be of different natures. Some map directly to real-world, measurable counterparts and some are simulation-specific with no direct real-life counterpart. Some of the former can be extracted from domain-specific knowledge (either experimental or theoretical). Others are design-specific.

A model's behavior space is expected to overlap somewhat with the behavior space of the referent system. Achieving that requires that model parameters be appropriately tuned (adjusted) to represent desired real-life situations. Each real-life situation has measurable properties (phenotypic attributes), which define its unique characteristics (phenotype). Each simulated situation is similarly characterized by its unique simulation parameters.

In real-life situations, a causal relationship exists between generative mechanisms and measured properties. A similar mapping exists for hepatocytes simulations. We follow an axiom that in many cases a mapping exists between the space of selected, measured properties and the space of simulation parameter values. Figure 2.10 illustrates this axiom: three different real-life situations are shown, two of which are closer together in the space of measurable properties. The arrangement of simulated phenomena relative

to the arrangement of simulation parameters may differ from that of real life, even though their relative distances to each other are more or less similar. Nevertheless, as a first approximation, we assume that the relationship between a new phenomenon and its acceptable simulation parameters can be approximated from its position relative to acceptable, previously simulated situations.

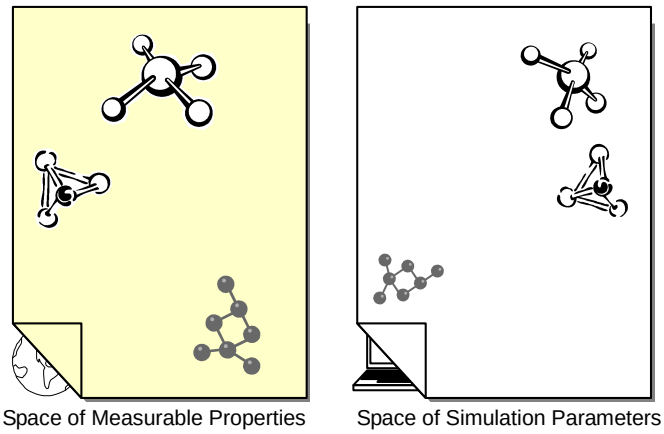

Figure 2.10 System *situations* with related, measurable properties and generating mechanisms are expected to have similarly related, *in silico* properties and generating mechanisms.

The PEAF Algorithm

In this section we present an algorithm that uses Fuzzy clustering to estimate simulation parameters for a new situation given the tuned parameters of several, previously validated situations. A Fuzzy classifier provides a measure of the degree to which a pattern fits within a class. There are several techniques for Fuzzy pattern recognition. In this work, we use a Fuzzy pattern recognition technique introduced in (56): the Fuzzy c-Means iterative algorithm.

The inputs of the Fuzzy c-Means algorithm are: 1) the set of n data points to be clustered, 2) number of clusters c, and 3) a parameter m known as the Fuzzy exponent.

As recommended by (57) we always set m = 2. The output of Fuzzy c-Means algorithm, U, is a c-by-n matrix, containing the values of the membership functions of the fuzzy clusters.

In general, for a data set, S, containing n situations S = {c_1, c_2,…, c_n}, the following PEAF algorithm is proposed to estimate the simulation parameters of a new situation, cn+1:

Step 1. Let q = n, and Snew= {c_1 , c_2… c_n , c_{n+1}}.

Step 2. If q = 1 go to step 4. Else, classify Snew into q clusters using Fuzzy c-Means algorithm.

Step 3. If c_{n+1} is not in the same group with at least another member then decrease q to q-1. Repeat steps 2 and 3.

Else, let G-value be the number of groupmates of c_{n+1}. Go to step 4.

Step 4. Call the q groups G_1, G_2 … G_q where $c_{n+1} \in G_1$. Let μk be the membership degree of c_{n+1} to Gk. Estimate the simulation parameters of c_{n+1} as:

$$\vec{P}_X = \sum_{k=1}^{q} \mu_k \cdot \vec{P}_{Gk} \tag{2.9}$$

where P_{Gk} is the weighted average parameter vector of all the members of group k:

$$\vec{P}_{Gk} = \frac{\sum_{j=1}^{m} \mu_j \cdot \vec{P}_j}{\sum_{j=1}^{m} \mu_j} \tag{2.10}$$

The accuracy or usefulness of the resulting estimates depends, of course, on how many situations similar to c_{n+1} exist in the data set, i.e. the higher the G-value, the better the

accuracy.

Estimating the Parameters of an Agent-Based Model

In this section we show how the PEAF algorithm can be used to estimate the simulation parameters of an agent-based model in order to make predictions. In this model, a situation is characterized by hepatocytes behavior in the presence of a particular compound. ISH2 is an agent-based model of hepatocytes (Section 2.2). The cells are simulated on a 2D grid; it maps to the culture dish. When hepatocytes are exposed to different simulated drug compounds, they metabolize and eliminate them, as *in vivo*. Consequently, simulation parameter values that are sensitive to physicochemical properties (PCPs) need to be different for each drug. The goal is to estimate the PCP-sensitive parameter values to enable simulating the metabolic and transport properties of a new drug given the parameter values similarly used and validated for several previously studied drugs.

To demonstrate, reconsider the four compounds shown in Table 2.4. The following PCPs were considered: molecular weight, logP, hydrogen bond donor count, hydrogen bond acceptor count, rotatable bond count, tautomer count, pKa, TPSA, volume, GPCR ligand, ion channel modulator, kinase inhibitor, and nuclear receptor ligand. The classification results for the PCPs of the four compounds clustered to two and three classes using the Fuzzy c-Means algorithm based on their PCPs are shown in Table 2.3. The classification results show that when divided into two groups, taurocholate, enkephalin, and methotrexate have more membership in the same group while salicylate belongs primarily to another. However, when divided to three groups, taurocholate and methotrexate have membership in the same group, whereas enkephalin and salicylate

belong primarily to different groups.

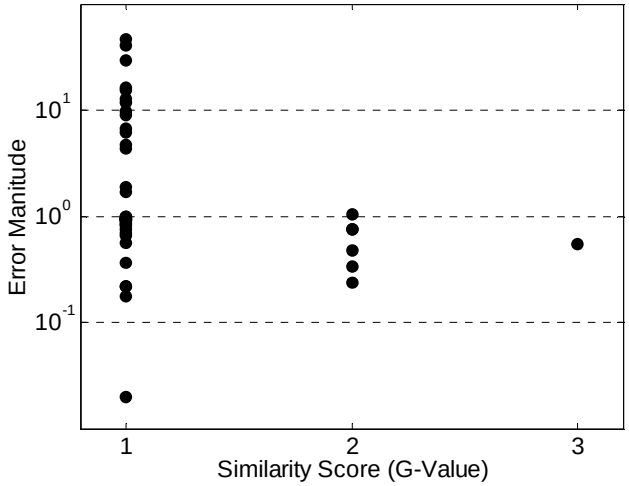

Figure 2.11 Prediction error decreases as G-value increases

Consider this task: based on the information in Table 2.3, we want to estimate the PCP-sensitive parameter values of enkephalin given the corresponding parameters of the other three compounds. There was no point in clustering the four drugs to four clusters, so we started with three. When c = 3, the Fuzzy c-Means algorithm provides no useful information about similarity of enkephalin to others: no other compound was in the same group with enkephalin (it however tells us about the dissimilarity of enkephalin to others). Thus, we took an additional step and clustered the compounds to two groups. When c = 2, enkephalin has two other groupmates. In that case, the best guess is that the PCP-sensitive parameter values for enkephalin are closer to those of its groupmates, taurocholate and methotrexate, than to salicylate. An intuitive way to estimate a parameter vector value for enkephalin is:

$$\hat{P}_{Enkeph.} = 0.8471\frac{(0.9264P_{Taur.} + 0.7692P_{Meth.})}{(0.9264 + 0.7692)} + 0.1529(0.9862P_{Sal.})$$

where P_x is the simulation parameter vector of compound x. In this example the *G-value*

is 2 because two compounds ended up within the cluster with enkephalin.

2.3.3 Results

The PEAF algorithm was used to iteratively predict the *clearance* of the 50 drugs (listed in

Table 2.6) in a leave-one-out process. Figure 2.11 shows that relative prediction error decreases as the *G-value* increases: the predictions are more accurate for compounds with more members in their clusters. Figure 2.12 shows the distribution of the prediction error. Compounds with *G-value* greater than 1 are located close to zero.

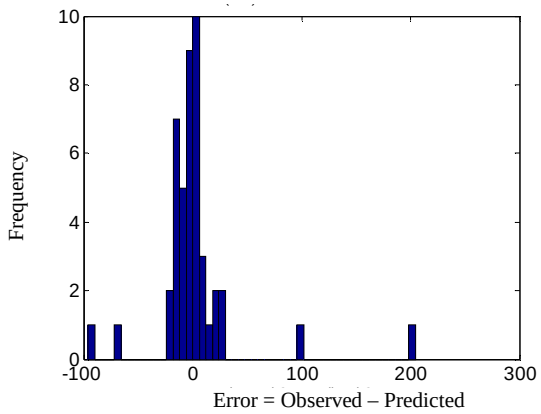

Figure 2.12 The Distribution of prediction error

2.3.4 Conclusion and Discussion

In this section we proposed a simple algorithm, called PEAF, that is based on Fuzzy clustering to estimate the PCP-sensitive parameter values of agent-based models. When the model's parameter values are properly tuned, it is capable of mimicking its referent. The PEAF algorithm utilizes the Fuzzy c-Means (FCM) algorithm to cluster previously encountered *situations* based on their measurable properties. The algorithm works based

on the assumption that similarity in the space of measurable properties maps to the similarity in the parameter space of the simulation model. PEAF offers important advantages: 1) because Fuzzy c-Means algorithm is unsupervised, the problem of over-fitting to the training data is minimized. That is particularly important in cases with small data sets. 2) The PEAF algorithm has no parameters. 3) It is relatively easy to implement. 4) The algorithm calculates a similarity score (the *G-value*) which correlates with the accuracy of its estimates: the higher the score, the higher the expected precision. As a result, the algorithm can advise in advance on the accuracy of its predictions.

As a proof of concept, we utilized the PEAF algorithm to estimate the PCP-sensitive parameter values of the ISH2 to *predict* the behavior of the referent system when it is introduced to a new compound not previously encountered. Note that parameter prediction is a direct mapping from the space of PCPs to the ISH2 parameter space, whereas parameter tuning draws its information from the biological behavior space. The estimated parameter values were fed to the ISH2 to enable it to make predictions. The predictions were compared to the observed measurements. For the seven of fifty compounds with G-values > 1 the predictions correlate nicely with the observed values as shown in Figure 2.13 ($p < 0.05$, $R^2 = 0.68$). We expect that as the number of drugs in the database increases, the probability that a compound similar to the new one of interest will exist in the database will increase; as a result better predictions can be anticipated.

Table 2.6 The Clearance values of 50 drugs (58) and their predicted values. Compounds with G-value>1 are shown in Bold.

	Drug Name	Clearance ($\mu L/min/10^6$ cells) Observed	Predicted	G-value
1	Bromocriptine	37	7.039	1
2	Caffeine	3.3	103.8	1
3	Carbamazepine	2.0	99.02	1
4	Cimetidine	1.2 ± 0.4	0.124	1
5	Cyclosporin A	3.5 ± 1.5	18.36	1
6	Diazepam	0.3	13.46	1
7	**Ethinylestradiol**	**7 ± 2.0**	**9.814**	**2**
8	Famotidine	< 1	17.38	1
9	Isradipine	18	6.214	1
10	Lorazepam	1.0	17.37	1
11	Midazolam	14 ± 8.0	1.051	1
12	Nifedipine	5.6 ± 1.5	13.52	1
13	Nitrendipine	7.4 ± 3.5	5.226	1
14	Omeprazole	1.7	47.68	1
15	Prazosin	2.3 ± 1.7	15.49	1
16	Propofol	107 ± 26	9.189	1
17	Ritonavir	2.1 ± 3.0	1.890	1
18	Temazepam	2.0	0.043	1
19	Triazolam	1.0	14.18	1
20	Zileuton	2.1 ± 1.8	2.521	1
21	Acebutolol	1.8 ± 1.5	16.13	1
22	Atenolol	< 1	0.815	1
23	Bepridil	2.0	18.97	1
24	Betaxolol	2.5 ± 1.0	17.19	1
25	Bisoprolol	1.6 ± 1.4	14.98	1
26	Carvedilol	35 ± 11	9.964	1
27	**Chlorpheniramine**	**2.8 ± 1.3**	**5.333**	**2**
28	Clozapine	6.0	1.846	1
29	Codeine	23	1.878	1
30	Desipramine	3.0	7.335	1
31	Dextromethorphan	7.6 ± 8.1	7.371	1
32	Diltiazem	9.0 ± 0.5	0.066	1
33	**Diphenhydramine**	**6.0**	**10.25**	**2**
34	**Doxepin**	**13**	**6.090**	**2**
35	Fluoxetine	1.0	13.16	1
36	Granisetron	9.0 ± 8.7	2.146	1
37	**Imipramine**	**8.0 ± 2.5**	**10.33**	**2**
38	Metoprolol	7.0 ± 2.9	12.40	1
39	Morphine	24	0.354	1
40	Nadolol	< 1	21.78	1

	Drug Name	Clearance ($\mu L/min/10^6$ cells) Observed	Predicted	G-value
41	Naloxone	216	12.35	1
42	Ondansetron	1.4 ± 0.5	1.202	1
43	**Pindolol**	**2.8 ± 1.0**	**0.662**	**2**
44	Pirenzepine	< 1	1.167	1
45	Propranolol	10 ± 0.5	7.073	1
46	Ranitidine	1.0 ± 0.0	0.700	1
47	Scopolamine	7.0	0.083	1
48	Triprolidine	4.3 ± 3.3	6.997	1
49	**Verapamil**	**18 ± 12**	**26.97**	**3**
50	Cetirizine	< 1	28.75	1

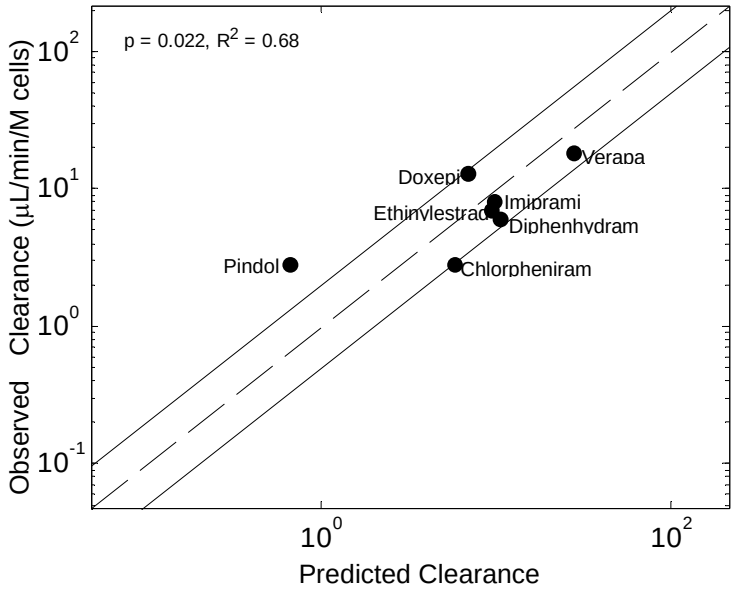

Figure 2.13 Predicted versus observed Clearance values for compounds having G-value > 1. Dotted line is the Identity line. 2-fold error boundaries are also shown (solid lines).

2.4 Prediction of In Vitro Drug Clearance Using In Silico Hepatocytes (ISH3) and Quantitative Structure to Micromechanism Relationships

2.4.1 Introduction

In this section we present an improved ISH model, called ISH3. It consists of several autonomous micromechanisms (Figure 2.14). In previous sections, ISHs were parameterized in part by tuning the parameter values to obtain the desired behavior. However in this section we do not adjust the parameters. Instead we estimate the parameters individually for each micromechanism by directly mapping the mechanisms to physicochemical properties.

2.4.2 Methods

Model Components

We used an agent-based method to construct the model (similar to the models described in sections 2.1, 2.2, and 2.3). Again we represent hepatocytes using fixed composite agents placed in a 2D grid where mobile objects representing solute can interact with them stochastically.

As in previous sections, to avoid confusion and clearly distinguish *in vitro* components and features from their *in silico* counterparts, such as a “hepatocyte,” a “solute,” or “excreted,” we use small caps when referring to the *in silico* system.

As shown in Figure 2.3, the incubation medium is modeled by a two dimensional

square grid space in which HEPATOCYTES and SOLUTES can be placed in separate grids to interact with each other.

- *DRUGS* (SOLUTES) are independent, mobile objects that move stochastically, governed by the flow of the incubation medium. During an experiment the event histories of SOLUTES (and other objects) can be tracked, such as SOLUTE that has been TRANSPORTED out of a CELL, or that has moved into a CELL.

- *HEPATOCYTES* are autonomous agents. Each is constructed from objects that map hepatocyte components and the environment: factors that can nonspecifically bind drug, enzymes and uptake/efflux transporters. The subcomponents interact with a DRUG according to four micromechanisms shown in Figure 2.14. Details follow.

- A *BINDER* is an object within a HEPATOCYTE that can bind to a free nearby SOLUTE and hold onto it for a specified number of simulation cycles (Figure 2.14A). Three parameters control the behavior of a binder:

Binding probability: the probability that the BINDER binds to a nearby DRUG.

Binding period: specifies how many simulation cycles, the BINDER holds the DRUG.

Release probability: the probability that the BINDER releases the DRUG after the binding period is over.

- An *ENZYME* is a specialized form of binder. It can METABOLIZE a bound SOLUTE by replacing it, following the binding period, with a METABOLITE and destroying the replaced SOLUTE. Currently we ignore the METABOLITE and simply destroy the SOLUTE (Figure 2.14B). In addition to the BINDER'S parameters, an ENZYME has an additional

parameter:

Metabolism probability: the probability that the bound SOLUTE is metabolized after the binding period has ended.

- *TRANSPORTERS* belong to a subclass of binders. They can bind with free SOLUTE that is either inside or outside, and transport them to the opposite side of the CELL MEMBRANE, independent of the local SOLUTE density (Figure 2.14C). When needed, TRANSPORTERS can be subdivided further into specialized forms.

In Silico Events

Once a free DRUG encounters a HEPATOCYTE, three randomized events can happen:

1. Passive movement into the CELL (Figure 2.14D): Two parameters, *Membrane_Cross-In_Probability* and *Cell_Capacity*, determine when a free SOLUTE that has encountered a CELL may enter it. The former, which is governed by solute properties, is the probability that the SOLUTE enters the CELL passively. The latter determines the number of objects a CELL can accommodate. Each unbound INTRACELLULAR SOLUTE may also partition out with a probability of *Membrane_Cross-Out_Probability.*

2. Active transport into the CELL: the DRUG can bind to an uptake TRANSPORTER with probability of *Uptake_Transport_probability*. If recognized, it is transported into the CELL. We assume that TRANSPORTERS are placed randomly within a CELL MEMBRANE.

3. Nonspecific binding to the MEMBRANE: the DRUG can bind to a nonspecific BINDER located on the outer side of the MEMBRANE.

Once a DRUG is within the CYTOSOL one or more of the following events (in a randomized order) occurs:

1. Passive movement out of the CELL through basal MEMBRANE according to *Basal_Membrane_Cross-out_Probability.*

2. Passive movement out of the CELL through the apical MEMBRANE according to the *Apical_Membrane_Cross-out_Probability*

3. Binding to a nonspecific BINDER. Once bound, it remains attached to the BINDER for the binding period of the BINDER-SOLUTE pair, after which the DRUG might be released according to the release probability.

4. Active transport out of the CELL by binding to an apical or basal efflux TRANSPORTER according to the TRANSPORTER-SOLUTE transport probability.

5. Binding to an ENZYME according to the ENZYME-SOLUTE binding probability. After the binding period has ended it might be metabolized according to the specified *Metabolism probability o*f the ENZYME-SOLUTE pair. If not METABOLIZED, the SOLUTE is released back to the CYTOSOL according to its release probability.

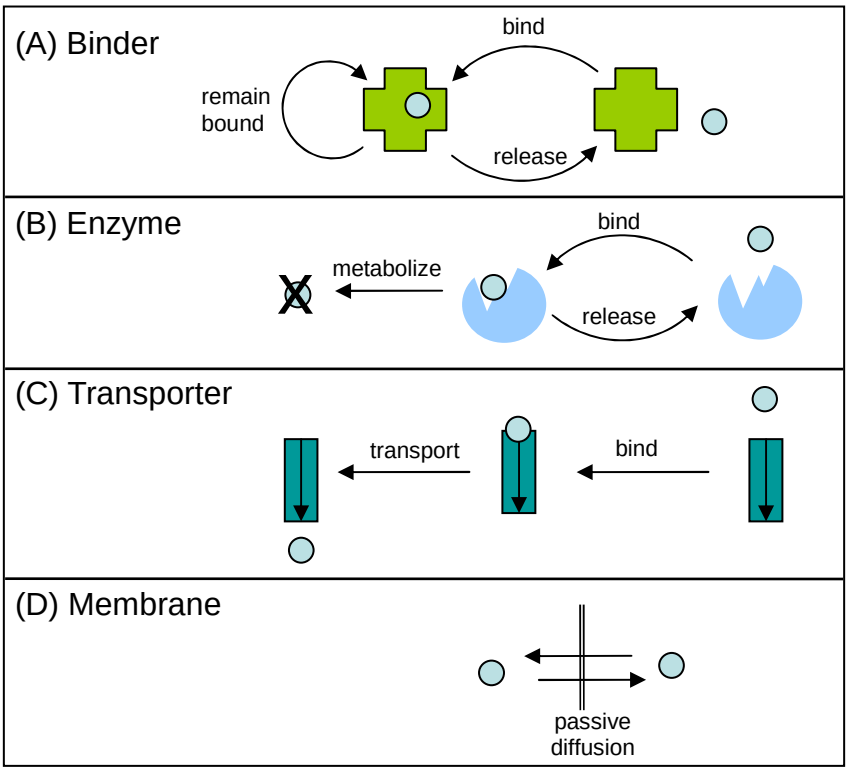

Figure 2.14 Schematic of four types of ISH3 micromechanisms. (A) A BINDER binds with a nearby SOLUTE, remains attached for a certain period (*Binding_Period*), and then releases the SOLUTE with a predetermined probability. If not released the process repeats. (B) ENZYMES bind and release SOLUTES similar to BINDERS, however there is a specified probability that one will METABOLIZE the SOLUTE at the end of the binding period. (C) TRANSPORTERS bind with free SOLUTES, remain attached for certain period, and then release them to the other side of the MEMBRANE. (D) A SOLUTE can cross MEMBRANES passively with a certain probability.

Parameter Estimation

For a given set of compounds, $\{X_i \mid i = 1,\ldots,n\}$, our goal is to estimate the parameter values of the micromechanisms (Figure 2.14). The execution of the parameterized micromechanisms causes emerged CLEARANCE properties that are mapped logically to the observed clearance properties. We specified reasonable mappings between physicochemical properties (PCP) of X_i's and the properties of micromechanisms. The relationships, defined as independent *axioms*, are based on published literature or expert knowledge. For example: Axiom #1: The more lipophilic *X* is, the more likely it is to be

bound to CYP2C9 (59, 60, 61) (in other words: LogP is positively correlated with CYP2C9 binding). Table 2.7 shows general relationships between several types of physicochemical properties and P450 enzyme binding. Table 2.8 shows relationships between physicochemical properties and membrane permeability.

After all axioms are specified, a *scoring* function was defined for each micromechanism as a weighted average of PCPs:

$$Score(\mathrm{M}) = w_1 \cdot PCP_1 + w_2 \cdot PCP_2 + \ldots + w_n \cdot PCP_n$$

where M is the desired micromechanism, w_i is a scalar which determines the degree to which PCP_i influences M. w_i is positive (negative) if PCP_i is positively (negatively) correlated with M.

The score functions were defined and calculated for all micromechanisms (Table 2.11). Normalized score values were used to obtain the desired parameter values. Details follow.

P450 Binding and Metabolism

We consider the six major forms of P450 isozymes that are mainly involved in the metabolism of xenobiotics in man (59): CYP3A4, CYP2D6, CYP2C9, CYP1A2, CYP2C19, and CYP2E1 with 34%, 19%, 16%, 8%, 8% and 4% involvement in drug oxidation, respectively (61).

Table 2.7. Influential PCPs on P450 enzyme binding. An upward arrow indicates positive, and a downward arrow indicates negative correlation.

property		CYP3A4 binding	CYP2D6 binding	CYP2C9 binding	CYP1A2 binding	CYP2C19 binding	CYP2E1 binding	P450 binding	P450 rate of metab-olism
Size	Molecular Weight	(↓) (60)	(↑) (61)			(↑) (61)	(↓) (59) (60) (61)		
Flexibility	Rotatable Bond Count			(↑) (60)	(↑) (60)	(↑) (60)		(↑) (60)	
Lipophilicity	logP	(↑) (59) (60)	(↑) (59)	(↑) (59) (60) (61)	(↑) (59)	(↑) (61)	(↑) (59)		
	logD							(↑) (60)	
Acidity	Hydrogen Bond Donor Count			(↑) (59) (60) (61)	(↑) (61)	(↓) (60) (61)	(↑) (61)		
	Hydrogen Bond Acceptor Count			(↑) (60)	(↑) (61)	(↑) (60)	(↑) (61)		
Ionization	pKa	(↑) (59)		(↑) (61)	(↑) (59)				
	Max fraction ionized at pH=7.4								(↑) (60)
	Ionization Potential (HOMO[1])	(↑) (61)			(↓) (61)				(↑) (60)
Polarity	Polar surface area (PSA)		(↑) (60)						
	Dipole moment				(↓) (61)				

[1] Highest occupied molecular orbit.

Table 2.8. Relationship between physicochemical properties and membrane permeability (membrane diffusion and transport). An upward arrow indicates positive, and a downward arrow indicates negative correlation.

	Membrane diffusion	**caco-2 permeability**	**ABCB1 substrate[1]**	**ABCB1 non-substrate[1]**
Molecular Weight	(↓) (58)		>400 (62)	<400 (62)
Number of N and O atoms			≥8 (62)	≤ 4 (62)
$logD_{7.4}$		(↑) (63)		
logP	(↑) (58)			
Acidic pKa			<4 (62)	
Basic pKa				<8 (62)
Ionization	(↓) (58)			
polar van der Waals' surface area (=TPSA)	(↑) (58)	(↓) (63)		

[1] "rule of fours" (62).

Table 2.9. Calculation of individual micromechanism scores as linear functions of PCPs.

CYP3A4 binding score	$BS_{3A4} = (-MW + 2 \cdot logP + IP + pKa_{acidic})/3$
CYP2D6 binding score	$BS_{2D6} = (MW + logP + TPSA + DM)/3$
CYP2C9 binding score	$BS_{2C9} = (RBC + 3 \cdot logP + 3 \cdot HBD + HBC - pKa^{1})/3$
CYP 1A2 binding score	$BS_{1A2} = (RBC + logP + HBD + HBA + pKa_{acidic} - IP - DM)/3$
CYP 2C19 binding score	$BS_{2C19} = (MW + RBC + logP + HBA - HBD)/3$
CYP 2E1 binding score	$BS_{2E1} = (-MW + 2 \cdot logP + HBD + HBA)/3$
P450 binding score	$BS_{P450} = (RBC + logD\ 7.5)/3$
p450 rate of metabolism score	$MS_{p450} = MFI + IP$
caco-2 permeability score	$PS_{caco-2} = logD\ 7.5 + TPSA$
P-gP substrate score	$SS_{P-gP} = (MW - 400)/400 + (ONC - 8)/8 + (4 - pKa_{acidic})/4$
P-gP nonsubstrate score	$NS_{P-gP} = (400 - MW)/400 + (4 - ONC)/4 + (8 - pKa_{basic})/8$
MEMBRANE diffusion score	$MDS = - MFI + TPSA + logP - MW$

[1] Acidic pKa was used.

Abbreviations: MW:Molecular Weight, RBC:Rotatable Bond Count, HBD:Hydrogen Bond Donor Count, HBC:Hydrogen Bond Acceptor Count, MFI: Max fraction ionized at pH=7.4, IP: Ionization Potential, TPSA: Topological Polar Surface Area, DM: Dipole Moment, ONC: O+N atom count.

Calculation of Overall CYP binding score (BS_{CYP})

BS_{CYP} was calculated as an average of the six CYP binding scores weighted by their percent involvement in drug oxidation:

$$BS_{CYP} = 0.34 \cdot BS_{3A4} + 0.19 \cdot BS_{2D6} + 0.16 \cdot BS_{2C9} + 0.08 \cdot BS_{1A2} + 0.08 \cdot BS_{2C19} + 0.04 \cdot BS_{2E1} + 0.11 \cdot BS_{P450}$$

Overall P-gP (apical efflux) transport score was calculated as:

$$TS_{P\text{-}gP} = (SS_{P\text{-}gP} - NS_{P\text{-}gP})/2$$

Calculation of Overall Basal Uptake Transport score (UTS_{basal})

Assuming $PS_{caco\text{-}2} \sim UTS_{basal} + TS_{P\text{-}gP}$ (for simplicity reasons hepatic apical uptake was neglected), we have:

$$UTS_{basal} = PS_{caco\text{-}2} - TS_{P\text{-}gP}$$

Calculation of Simulation Parameters from Scores

Parameter values for the i^{th} compound were calculated from above scores as follows:

$$pCypBind_i = 0.9\ \{BS_{CYP,i} - \min(BS_{CYP})\}/\{\max(BS_{CYP}) - \min(BS_{CYP})\} + 0.05$$

$$pCypMetabolize_i = 0.9\ \{MS_{p450,i} - \min(MS_{p450})\}/\{\max(MS_{p450}) - \min(MS_{p450})\} + 0.05$$

$$pUptakeTransport_i = 0.9\ \{UTS_{basal,i} - \min(UTS_{basal})\}/\{\max(UTS_{basal}) - \min(UTS_{basal})\} + 0.05$$

$$pEffluxTransport_i = 0.9\ \{TS_{P\text{-}gP,i} - \min(TS_{P\text{-}gP})\}/\{\max(TS_{P\text{-}gP}) - \min(TS_{P\text{-}gP})\} + 0.05$$

(both apical and basal MEMBRANES used the same TRANSPORT properties)

$$pCypRelease_i = 1 - \mathrm{sqrt}(pCypBind_i)$$

(The above equation was used to make the probability of remaining bound greater than the probability of binding).

$$pBasalCrossIn_i = 0.5\ \{TS_{P\text{-}gP,i} - \min(TS_{P\text{-}gP})\}/\{\max(TS_{P\text{-}gP}) - \min(TS_{P\text{-}gP})\} + 0.01$$

$pBasalCrossOut_i$ = 0.01 *pBasalCrossIn*

$pApicalCrossIn_i$ = 0.1 *pBasalCrossIn*

$pApicalCrossOut_i$ = 0.01 *pBasalCrossIn*

Other parameters were kept constant for all DRUGS:

$nCypEnzymesPerCell_i$ = 3

$nCypBindingCycles_i$ = 2

$nUptakeTransporterPerCell_i$ = 1

$nUptakeTransportCycles_i$ = 2

$nEffluxTransporterPerCell_i$ = 1

$nEffluxTransportCycles_i$ = 2

$nCellCapacity_i$ = 10

Crude In Silico Clearance Measurements

In vitro intrinsic clearance can be calculated using area under concentration curve and dose: CL_{int} = dose/AUC. We used a similar method to calculate *in silico* intrinsic clearance. The time course of fraction of SOLUTES in EXTRACELLULAR space during the simulation was recorded. Note that the area under this fraction curve ($AUFC_{in\ silico}$) maps to the area under concentration curve divided by the dose (AUC/dose). Consequently the *in silico* CL was calculated as[1]: $CL_{in\ silico}$ = DOSE/AUC = $1/AUFC_{in\ silico}$.

Calculation of CL_{int} Predictions

Predicted ISH3 CL values were calculated in a leave-one-out manner as follows: the i^{th} crude *in silico* CL ($CL_{crude,i}$) (Table 2.11) was taken out of the set of the 39 values, and

[1] Volume of distribution (V) is not defined *in silico*.

its predicted CL_i value was calculated as: ($scale_i$).($CL_{crude,i}$). $scale_i$ is chosen such that optimally transforms the remaining 38 CL values within the 95% confidence intervals.

For comparison, CL values were predicted using multiple regression (MR) as well. MR predictions were also calculated in a leave-one-out manner. Each time one compound was taken out of the data set, the regression parameter values were calculated for the remaining 72 compounds. Independent variables were the PCPs (Table 2.7).

2.4.3 Results

Table 2.10 shows reported human, *in vitro*, hepatic intrinsic clearance values of 73 compounds (58, 64, 65). For 39, of those compounds more than two reported values were available (n>2), for which we were able to construct 95% confidence intervals. The Table also shows the class of the drugs according to the Biopharmaceutics Classification System (BCS) (66). 17 compounds belong to Class 1 (high permeability, high solubility), 6 compounds belong to Class 2 (high permeability, low solubility), 6 compounds belong to Class 3 (low permeability, high solubility), and only 1 compound belongs to Class 4 (low permeability, low solubility). BCS classifications were not available for other compounds.

Table 2.11 shows the crude average *in silico* clearance measurements for 10 Monte Carlo simulation runs. The simulation parameters for each drug were estimated as described in Methods (Section 2.4.2). No parameter tuning was carried out. Predicted clearance values are shown in Table 2.12. 77% of the predicted clearance values were within 95% confidence interval of the *in vitro* clearance values. For comparison, predicted values obtained by multiple regression are also shown; 33% of those values fall

within the 95% confidence interval. The Table also shows the BCS class of drugs. If we limit attention to the individual BCS classes we can see that the predictions for 12 out of 13 (92%) class 1 compounds, 1 out of 2 (50%) class 2 compounds, and 6 out 6 (100%) class 3 compounds are acceptable.

2.4.4 Discussion

In this section we presented a 2-step method to estimate parameterizations for ISH3 micromechanisms in order to predict *in vitro* hepatic clearance of drugs. First, we specified linear mappings between physicochemical properties (PCPs) and properties of ISH3 micromechanisms. The mappings were based on published quantitative structure-activity relationships found in the literature. They transformed a compound from its space of physicochemical properties to an *m*-dimensional score space, where *m* is the number of micromechanism parameters. Next, the score values were used to estimate the parameters of micromechanisms. In doing so we assumed that parameter values have positive correlation with score values, such that higher scores result in larger parameter values and vice versa.

The linear mappings and the parameter estimation method presented herein impose biological constrains and requirements on the space of software mechanisms (and parameterizations) that can cause the emergence of acceptable *in silico* clearance values. Doing so shrinks the mechanism space. A continuation of this process will lead to *in silico* micromechanisms and parameterizations that increasingly mimic their *in vitro* counterparts.

For comparison, we used multiple regression (MR) analysis to predict *in vitro*

clearance from PCPs. The accuracy of MR predictions was significantly less than that of ISH3 predictions. Sophisticated data mining (inductive) techniques such as SVM or ANN might provide more accurate predictions. However such methods do not provide insight into the underlying mechanisms that play roles in emergence of the phenomenon (*in vitro* hepatic clearance of drugs).

Table 2.10 Human *in vitro* hepatic intrinsic clearance of 73 compounds. We were able to construct 95% confidence intervals for only 39. Data is compiled from (58, 64, 65). BCS classes are reported in (66).

	Compound name	**BCS[1] class**	**Human CLint (uL/min/1e6 cells)** Average	SD	**number of samples (n)**	**95% confidence interval**	
1	Bromocriptine	n/a	37.00		1		
2	Caffeine	1	1.60	1.61	3	-2.40	5.60
3	Carbamazepine	2	2.00		1		
4	Cimetidine	3	1.20	0.40	9	0.89	1.51
5	Cyclosporin A	2	3.50	1.50	8	2.25	4.75
6	Diazepam	1	0.80	0.56	3	-0.58	2.18
7	Ethinyl estradiol	1	7.00	2.00	5	4.52	9.48
8	Famotidine	3	0.50	0.50	6	-0.02	1.02
9	Isradipine	n/a	18.00		1		
10	Lorazepam	n/a	0.52	0.41	3	-0.50	1.55
11	Midazolam	1	11.64	16.32	7	-3.45	26.73
12	Nifedipine	1	5.95	7.66	7	-1.13	13.03
13	Nitrendipine	n/a	7.98	12.52	9	-1.65	17.60
14	Omeprazole	n/a	1.70		2		
15	Prazosin	n/a	2.30	1.70	7	0.73	3.87
16	Propofol	n/a	107.00	26.00	5	74.72	139.28
17	Ritonavir	2	2.10	3.00	7	-0.67	4.87
18	Temazepam	n/a	2.00		2		
19	Triazolam	n/a	1.00		2		
20	Zileuton	n/a	2.10	1.80	16	1.14	3.06
21	Acebutolol	n/a	1.80	1.50	4	-0.59	4.19
22	Atenolol	3	0.50	0.50	5	-0.12	1.12
23	Bepridil	n/a	2.00		1		
24	Betaxolol	n/a	2.50	1.00	6	1.45	3.55
25	Bisoprolol	n/a	1.60	1.40	8	0.43	2.77
26	Carvedilol	2	35.00	11.00	5	21.34	48.66

	Compound name	BCS[1] class	Human CLint (uL/min/1e6 cells) Average	SD	number of samples (n)	95% confidence interval	
27	Chlorpheniramine	1	2.80	1.30	4	0.73	4.87
28	Clozapine	n/a	6.00		1		
29	Codeine	n/a	23.00		2		
30	Desipramine	1	7.00	5.66	2	-43.82	57.82
31	Dextromethorphan	n/a	7.60	8.10	81	5.81	9.39
32	Diltiazem	1	8.87	100.80	90	-12.24	29.99
33	Diphenhydramine	1	6.00		1		
34	Doxepin	1	13.00		1		
35	Fluoxetine	1	1.00		1		
36	Granisetron	n/a	9.00	8.70	4	-4.84	22.84
37	Imipramine	1	8.05	5.34	4	-0.45	16.55
38	Metoprolol	1	7.00	2.90	5	3.40	10.60
39	Morphine	n/a	24.00		2		
40	Nadolol	3	0.50	0.50	3	-0.74	1.74
41	Naloxone	n/a	86.90	111.95	4	-91.23	265.03
42	Ondansetron	n/a	1.40	0.50	5	0.78	2.02
43	Pindolol	n/a	2.80	1.00	4	1.21	4.39
44	Pirenzepine	n/a	0.50	0.50	4	-0.30	1.30
45	Propranolol	1	9.94	102.82	92	-11.36	31.23
46	Ranitidine	3	1.00	0.05	5	0.94	1.06
47	Scopolamine	n/a	7.00		1		
48	Triprolidine	n/a	4.30	3.30	4	-0.95	9.55
49	Verapamil	1	17.98	141.36	88	-11.98	47.93
50	Cetirizine	3	0.50	0.10	5	0.38	0.62
51	Tenoxicam	n/a	2.60		1		
52	Warfarin	2	1.10		1		
53	Tolbutamide	n/a	1.60		1		
54	Antipyrine	1	0.29	0.26	2	-2.07	2.64
55	Furosemide	3&4	0.00		1		
56	Theophylline	1	0.31	0.30	2	-2.36	2.98
57	Ibuprofen	2	4.20		1		
58	Terbutaline	n/a	0.00		1		
59	Oxazepam	n/a	1.20	1.13	2	-8.96	11.36
60	Sulpiride	n/a	0.00		1		
61	Sildenafil	n/a	5.20		1		
62	Methylprednisolone	n/a	9.70		1		
63	Chlorpromazine	2	11.00		1		
64	Prednisolone	1	9.70		1		
65	Chlorprothixene	n/a	14.00		1		

	Compound name	BCS[1] class	Human CLint (uL/min/1e6 cells) Average	SD	number of samples (n)	95% confidence interval	
66	Tolcapone	n/a	1.20		1		
67	Bosentan	n/a	0.20		1		
68	Mibefradil	n/a	0.90		1		
69	Nicardipine	n/a	7.30		1		
70	Mofarotene	n/a	2.00		1		
71	Felodipine	n/a	7.50		1		
72	Remikiren	n/a	19.50		1		
73	Nilvadipine	n/a	13.30		1		

[1] Biopharmaceutics Classification System: Class 1 - High Permeability, High Solubility; Class 2 - High Permeability, Low Solubility; Class 3 - Low Permeability, High Solubility; Class 4 - Low Permeability, Low Solubility.

Table 2.11 The crude (untreated) average *in silico* CL results from ISH3.

	Compound	*in silico* CL
1	Caffeine	0.17
2	Cimetidine	0.22
3	Cyclosporin A	0.28
4	Diazepam	0.26
5	Ethinylestradiol	0.25
6	Famotidine	0.18
7	Lorazepam	0.29
8	Midazolam	0.27
9	Nifedipine	0.27
10	Nitrendipine	0.29
11	Prazosin	0.17
12	Propofol	0.27
13	Ritonavir	0.31
14	Zileuton	0.20
15	Acebutolol	0.19
16	Atenolol	0.16
17	Betaxolol	0.20
18	Bisoprolol	0.19
19	Carvedilol	0.23
20	Chlorpheniramine	0.24
21	Desipramine	0.19
22	Dextromethorphan	0.25
23	Diltiazem	0.19
24	Granisetron	0.17
25	Imipramine	0.23
26	Metoprolol	0.15
27	Nadolol	0.14
28	Naloxone	0.15
29	Ondansetron	0.19
30	Pindolol	0.17
31	Pirenzepine	0.08
32	Propranolol	0.19
33	Ranitidine	0.21
34	Triprolidine	0.26
35	Verapamil	0.29
36	Cetirizine	0.14
37	Antipyrine	0.23
38	Theophylline	0.10
39	Oxazepam	0.28

Table 2.12 ISH3 prediction results compared with multiple regression predictions. 77% of ISH3 predicted values lie within the corresponding 95% confidence interval. But 33% of multiple regression predicted values lie within the corresponding 95% confidence interval. The mean error magnitudes are 8.7 and 14, respectively.

	Compound	BCS class	predicted CL by ISH3	within 95% confidence?	predicted CL by multiple regression	within 95% confidence?
1	Caffeine	1	0.84	Y	-12.79	n
2	Cimetidine	3	1.12	Y	12.54	n
3	Cyclosporin A		1.41	n	33.22	n
4	Diazepam	1	1.30	Y	9.70	n
5	Ethinylestradiol		1.26	n	31.62	n
6	Famotidine	3	0.91	Y	8.34	n
7	Lorazepam		1.45	Y	15.47	n
8	Midazolam	1	1.37	Y	11.51	Y
9	Nifedipine	1	1.33	Y	4.86	Y
10	Nitrendipine		1.43	Y	6.39	Y
11	Prazosin		0.86	Y	21.43	n
12	Propofol		1.35	n	-1.82	n
13	Ritonavir	2	1.57	Y	12.49	n
14	Zileuton		0.98	n	19.86	n
15	Acebutolol		0.93	Y	-0.32	Y
16	Atenolol	3	0.81	Y	-1.17	n
17	Betaxolol		0.98	n	3.89	n
18	Bisoprolol		0.96	Y	0.69	Y
19	Carvedilol	2	1.14	n	12.91	n
20	Chlorpheniramine	1	1.18	Y	-3.33	n
21	Desipramine	1	0.96	Y	15.41	Y
22	Dextromethorphan		1.26	n	10.12	n
23	Diltiazem	1	0.96	Y	12.58	Y
24	Granisetron		0.84	Y	8.59	Y
25	Imipramine	1	1.14	Y	11.11	Y
26	Metoprolol	1	0.76	n	56.39	n
27	Nadolol	3	0.71	Y	2.75	n
28	Naloxone		0.73	Y	-0.51	Y
29	Ondansetron		0.97	Y	13.53	n
30	Pindolol		0.85	n	13.51	n
31	Pirenzepine		0.38	Y	5.19	n
32	Propranolol	1	0.95	Y	11.36	Y
33	Ranitidine	3	1.05	Y	0.33	n
34	Triprolidine		1.29	Y	9.96	n
35	Verapamil	1	1.47	Y	-7.95	Y

	Compound	BCS class	predicted CL by ISH3	within 95% confidence?	predicted CL by multiple regression	within 95% confidence?
36	Cetirizine	3	0.71	Y	4.07	n
37	Antipyrine	1	1.16	Y	-1.49	Y
38	Theophylline	1	0.52	Y	5.95	n
39	Oxazepam		1.39	Y	14.40	n
Percent within 95% confidence intervals			77%		33%	

3 Liver Zonation

Hepatic zonation is apparent periportal to perivenous attribute gradients within lobules. No concrete, causal, mechanistic theory is available to explain how different hepatic zonation patterns of P450 isozyme levels and hepatotoxicity emerge following dosing with different compounds. Zonation may have roots in an evolutionarily important hepatic role: defend the organism against the potentially damaging consequences of orally absorbed toxins. During its evolution, a hepatocyte, learned to detect and clear xenobiotics. Based on the fact that evolution appears to favor the development of species that utilize and retain energy more efficiently (67), we can theorize that during evolution, in the context of their multiple roles, hepatocytes, possibly other liver cells as well, have striven to achieve a close-to-optimal strategy for clearing such compounds. The theory may have a real time counterpart: upon extended exposure to a toxin, hepatocytes can revise their clearance strategies in real time to avoid or minimize risk of extrahepatic tissue damage, and that adjustment can be location dependent and influenced by the adjustments made by other hepatocytes. In this Chapter we present computational models designed to have features similar to relevant

hepatic features at specific levels of abstraction. The models include a simple two-player game, a multi-player game, three multi-agent models in which agents, mapping to sinusoidal segments, learned from experience to optimize their clearance strategies.

In Section 3.1 (adapted from (68) and (69) with minor revisions), we use game theory and reinforcement learning, to create and analyze generalized agent-based and compartmental models of hepatic toxin elimination processes to explore plausible causes of hepatic functional zonation. We considered a general situation in which a group of protective agents (analogous to liver cells) cooperate and self-organize their efforts to minimize optimally the negative effects of toxin intrusions. Following a totally different approach, we constructed a physiologically based model of a two-zoned liver to study the physiological consequences of zonation. The results of the two models support the hypothesis that liver zonation might be a consequence of an optimal strategy for toxin clearance.

In Section 3.2, we used the synthetic method of modeling and simulation to discover, explore, and experimentally challenge concrete mechanisms that show how and why biomimetic zonation patterns emerge and change within agent-based analogues, expecting that those mechanisms may have counterparts in rats. Mobile objects map to compounds. One analogue is comprised of 460 identical, quasi-autonomous functional units called sinusoidal segments (SSs). SSs detect and respond to compound-generated response signals and the local level of an endogenous gradient. Each SS adapts to new information with the objective of improving efficiency. Upon compound exposure, analogues developed a variety of patterns that were strikingly similar to those reported in

the literature. A degree of quantitative validation was achieved against data on hepatic zonation of CYP1A2 mRNA expression caused by three different doses of TCDD (2,3,7,8-tetracholorodibenzo-p-dioxone).

3.1 Multi-Agent Based Modeling of Liver Detoxification: Understanding the Role of Liver Zonation in Toxin Elimination[1]

3.1.1 Introduction

The liver performs a wide range of functions including detoxification of blood-borne compounds, and so doing protects the body. A human cannot live more than 24 hours without liver. Hepatocytes, the parenchymal cells of the liver, cooperate with each other to detoxify xenobiotics by metabolizing them to less toxic compounds. Over the course of their evolution they have learned to do so in an effective and optimal way. Hepatocytes express heterogeneous, location-dependent enzyme and transporter activities to facilitate detoxification, apparently following an intrinsic agenda, the principles of which are not fully understood. This phenomenon is known as liver zonation (70). To gain insight into those processes, we constructed and analyzed a generalized problem of cooperative agents protecting a commonwealth from harmful intruders. The agents are assumed to have incomplete information about each other and cannot form coalitions.

3.1.2 Biology Background

The liver is a complex biochemical factory which synthesizes, modifies, and metabolizes thousands of substances daily and provides the body with essential substances such as proteins and fats. The liver is also responsible for eliminating toxins and xenobiotics (including drugs) that find their way into blood. The rate of elimination,

[1] This Section is adapted from (68) and (69) with minor revisions

known as *hepatic clearance*, is different for each compound. Histologically, the liver is divided into lobules. Lobules consist of hepatocytes arranged in a roughly cylindrical or spherical shape. The central vein (CV), through which blood exits, is at the center. At the periphery are portal vein (PV) triads. A lobule is often described as being organized into three zones: periportal (upstream or zone 1), which encircles the portal tracts where blood enters, middle (zone 2), and perivenous (downstream or zone 3), which is poorly oxygenated and located around central vein (Figure 3.1). Oxygenated blood enters upstream, passes through the mid-zone, and exits downstream. Because of this spatial topology, different liver cells may not have the same exposure to incoming resources and compounds. For example, nutrients (e.g. oxygen) are more available to upstream than to the downstream cells.

Hepatocytes, although genetically identical, exhibit heterogeneous enzyme and transporter activities depending on their location within the lobule. For example, under normal conditions, hepatocytes located downstream express more enzymes for xenobiotic metabolism than do upstream hepatocytes. An obvious question is: why?

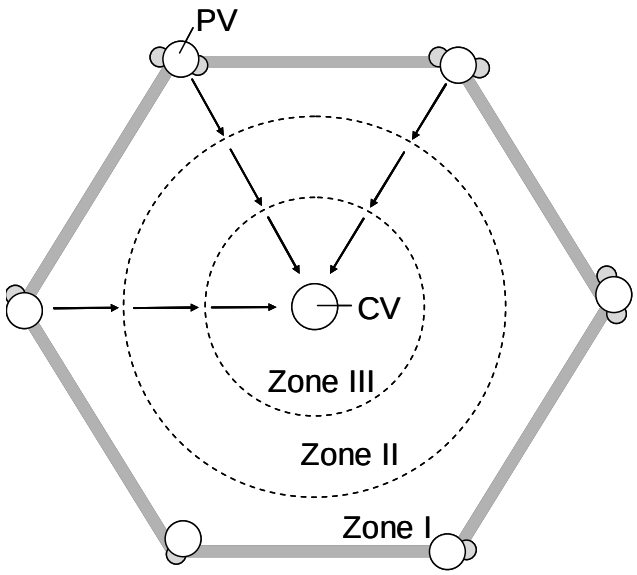

Figure 3.1. Cross section of a liver lobule. PV: portal triads, CV: central vein, arrows: direction of blood flow. Oxygenated blood enters the lobule from portal veins, and exits from central vein. Usually a liver lobule is described as being divided into three functional zones.

To represent hepatic metabolic zonation in physiologically based, pharmacokinetic models, researchers (71, 72) usually divide the liver into compartments, each representing a different intrahepatic zone. Christoffels et al. (73) presents a "mechanistic model" which proposes that zonation is induced by portocentral signal gradients. Using a different approach, we present an agent-based model of zonation.

3.1.3 Methods

All existing models of liver zonation are top-down models. They fail to hypothesize elementary mechanisms that motivate the collective behaviors of liver cells.

To understand the costs and benefits that may be associated with liver zonation, we began by using the game theoretic model as shown in Figure 3.2: *n* agents in a sequence are protecting their commonwealth (all extrahepatic tissues) against intruders. Agents

are limited in their ability to eliminate intruders. The goal of each is to minimize potential damage caused by the intruders while minimizing resource consumption. What is the optimal elimination strategy for agent *i*? Obviously, an optimal strategy for agent *i* depends on the strategies of other agents who have the same goal.

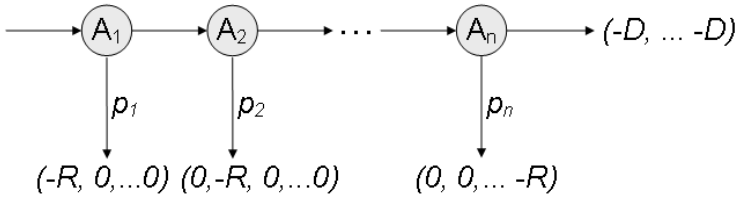

Figure 3.2. The game-theoretic model of the system. Agents either ignore or eliminate intruders. If Agent *i* (A_i) eliminates an intruder, it pays the cost of resource consumption (*R*). If all agents ignore an intruder, all must pay cost *D*, which is a consequence of damage caused by the intruder to the commonwealth. p_i is the elimination strategy of agent *i*: it is the probability that A_i eliminates an incoming intruder. A_i's ability to eliminate is limited by maxEi ($0 \leq \text{maxEi} \leq 1$).

Each agent has two options: eliminate or ignore an intruder. The immediate cost for elimination is resource consumption, denoted by *R*. Ignoring an intruder does not constitute an immediate cost; however, when all agents ignore an intruder (or it escapes for whatever reason), then all must pay the cost associated with any damage caused to the commonwealth (denoted by *D*). It is assumed that a signal informs agents of the damage cost at the end of each round of play. Agent *i* cannot eliminate more than $maxE_i$ fraction of incoming intruders even if it expends maximum elimination effort $(0 \leq maxEi \leq 1)$. The elimination strategy of agent *i*, p_i, is the probability that it eliminates an incoming intruder. Agents who see intruders earlier, are called upstream agents; the others are called downstream agents.

To analyze the game's equilibrium, we first specified that there are only two agents. Their cost functions are calculated as follows:

$e_1 = 1 - maxE_1 \cdot p_1$

$e_2 = (1 - maxE_2 \cdot p_2) \cdot e_1$

where e_1 and e_2 are the fraction of intruders that escape from Agents 1 and 2. Expected average costs of damage to each agent will be:

$ADC = e_2 \cdot D = (1 - maxE_2 \cdot p_2) \cdot e_1 \cdot D$

Expected average costs of resource consumption are:

$ARC_1 = maxE_1 \cdot p_1 \cdot R$

$ARC_2 = maxE_2 \cdot p_2 \cdot e_1 \cdot R$

Total expected costs are:

$$<Cost_1> = ARC_1 + ADC = maxE_1 \cdot p_1 \cdot R + (1 - maxE_2 \cdot p_2)(1 - maxE_1 \cdot p_1) \cdot D$$

$$<Cost_2> = ARC_2 + ADC = maxE_2 \cdot p_2 \cdot (1 - maxE_1 \cdot p_1) \cdot R + (1 - maxE_2 \cdot p_2)(1 - maxE_1 \cdot p_1) \cdot D$$

$$<Cost_1> = maxE_1 \cdot p_1 \cdot (R - D) + \{1 - maxE_2 \cdot p_2 \cdot (1 - maxE_1 \cdot p_1)\} \cdot D \quad (3.1)$$

$$<Cost_2> = maxE_2 \cdot p_2 \cdot (1 - maxE_1 \cdot p_1)(R - D) + (1 - maxE_1 \cdot p_1) \cdot D \quad (3.2)$$

Figure 3.3 shows the above cost functions for different values of D. At any given location in the strategy space, agents have a preferred direction of movement to reduce their costs. For the two-player game, the direction can be described as a vector field based on the gradients of the above two cost functions:

$$u = -\frac{\partial < Cost_1 >}{\partial p_1}, \quad v = -\frac{\partial < Cost_2 >}{\partial p_2}$$

The vector field is shown in Figure 3.4 for D/R = 0.6, 1.1, 1.7 and 2.3. Figure 3.5 shows

the vector field for n = 3 and D/R = 0.6, 1.1, 1.7, 2.3, 3.7 and 10.0. It is easy to find the equilibrium of each game by inspecting its vector field. The figures show that the equilibrium changes as D/R increases. When damage is very small ($D/R < 1$), all agents ignore because it is not cost effective to eliminate intruders. When damage is moderate, only downstream agents expend elimination effort. When damage is large, middle agents cooperate with the downstream agents. When damage is large enough, all agents expend elimination effort.

The analysis can be extended to a general case of *n* players as follows:

$$e_i = (1 - maxE_i \cdot p_i) \cdot e_{i-1} \qquad i = 1, 2, \ldots, n \qquad e_o = 1$$

$$ADC = e_n \cdot D$$

$$ARC_i = e_{i-1} \cdot maxE_i \cdot p_i \cdot R$$

$$<Cost_i> = ARC_i + ADC$$

$$<Cost_i> = e_{i-1} \cdot maxE_i \cdot p_i \cdot R + e_n \cdot D \qquad (3.3)$$

where e_i is the fraction intruders that escape from agent *i*; *ADC* is the average damage cost to each agent; ARC_i is the average cost of resource consumption to agent *i*; and $<Cost_i>$ is the total expected cost (due to both actions) to agent *i*. The vector field can be calculated the same as for the two-player game, but it is infeasible to visualize and find equilibria. In general, analyzing equilibria of games involving three or more players is hard (74, 75).

The above analysis requires that all agents have a priori knowledge about other agents and the environment. All the actions available to other agents and all costs with

all combinations of actions are required to be known by all agents. Consequently, it does not offer a plausible mechanism through which autonomous agents can reach an optimal strategy.

We considered a more realistic situation in which agents do not have a priori information about their environment (including other agents). We used multi-agent simulation and enabled agents to learn from experience following a simple reinforcement learning rule. By keeping track of accumulated reward (and penalty), agents could be reinforced to learn an optimal clearance strategy. Their task was to maximize the long-term average reward per action.

The Q-learning algorithm (76), a well known reinforcement learning algorithm, has been shown to converge to an optimal decision policy. Q-learning has a solid foundation in the theory of Markov decision processes (77). It is easy to implement and has been used widely in both single-agent and multi-agent contexts (see (77) and (78) for examples and (79) for a review of other multi-agent learning techniques).

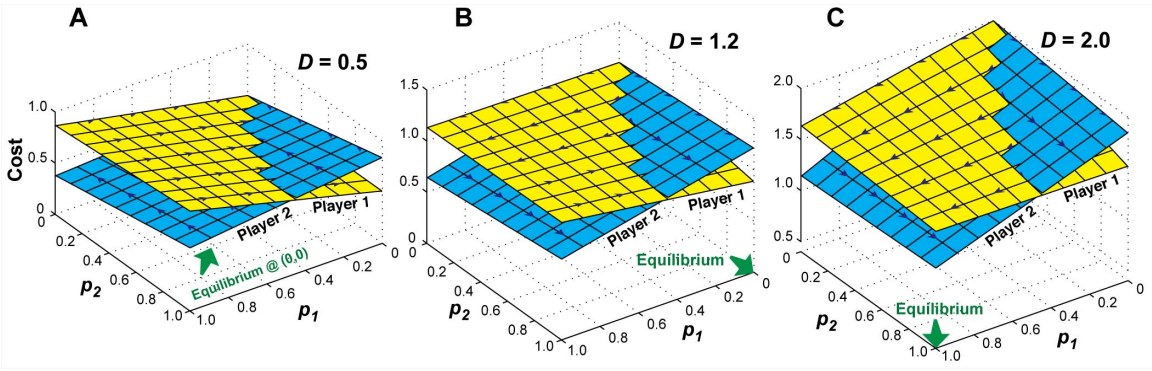

Figure 3.3 Two-player game cost functions versus players' strategies. Player's cost functions (equations (3.1) and (3.2)) are plotted against their clearance strategies (p_1 and p_2) for three different values of D. Arrows on the surfaces show players' preferred moving direction in order to reduce their cost function. A green arrow designates the game's equilibrium. Players' cost functions and the game equilibrium are influenced by toxin D value as shown. (A) When D is small (0.5), the game equilibrium is at (p_1 = 0,

$p_2 = 0$). Player 1's cost surface is steeper than that of player 2. As a result, player 2's contribution is less costly. (B) When D is larger, in this case $D = 1.2$, the game's equilibrium moves to ($p_1 = 0$, $p_2 = 1$): player 1 does nothing while player 2 expends maximum clearance effort. (C) Another equilibrium shift occurs when D is large enough. In this case, D must be at least 64 percent higher than the cost: both players expend maximum effort to clear toxins ($p_1 = 1$, $p_2 = 1$). For the results shown R=1.

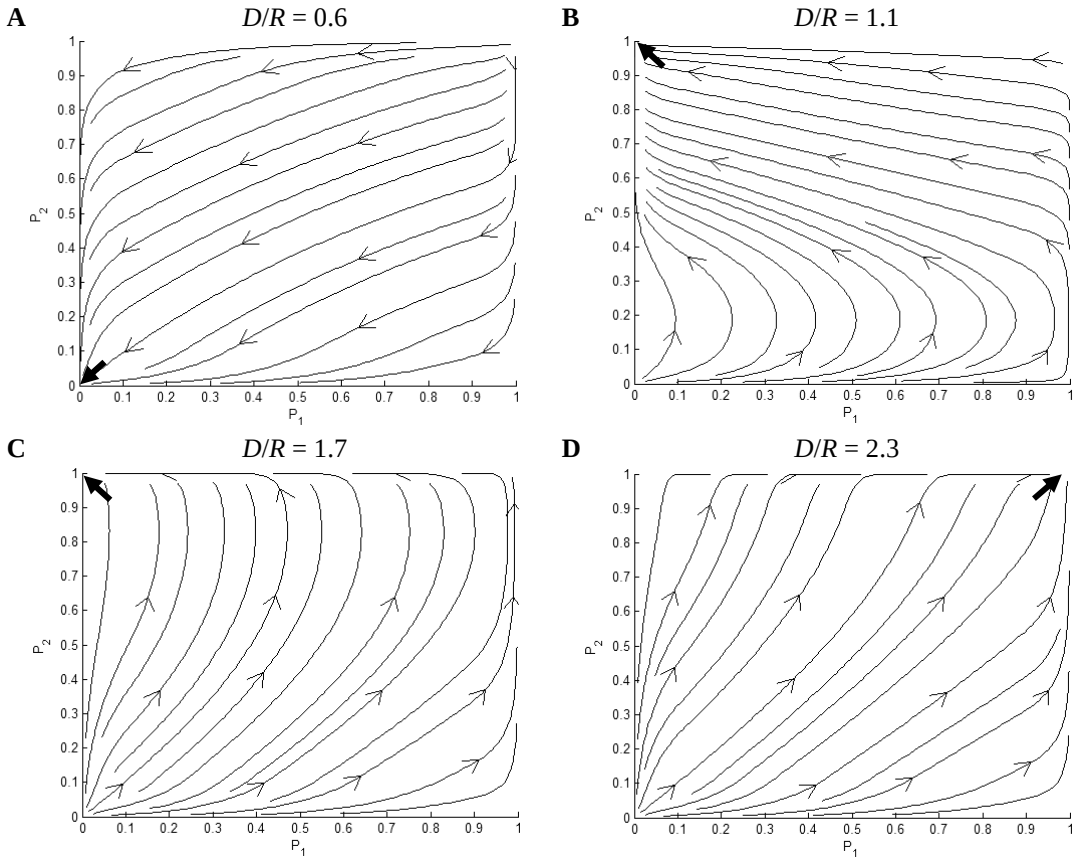

Figure 3.4. At any given non-equilibrium point in the strategy space, the agents have a preferred direction of movement in order to decrease their expected costs. The net direction towards which one moves in the strategy space depends on the slope of the cost functions at that point. Arrows show the net direction of movement when n = 2 and D/R ratio changes from 0.6 to 2.3. Each game's equilibrium is shown by a large arrow. (A) D/R is small. In this case, both agents ignore the intruders. (B, C) D/R is moderately large; the equilibrium is such that Agent 1 ignores (p1 = 0) but Agent 2 elimination effort is maximum (p2 = 1). (D) When D/R is large enough, the equilibrium changes to (p1 = 1, p2 = 1): the elimination effort of both agents is maximal.

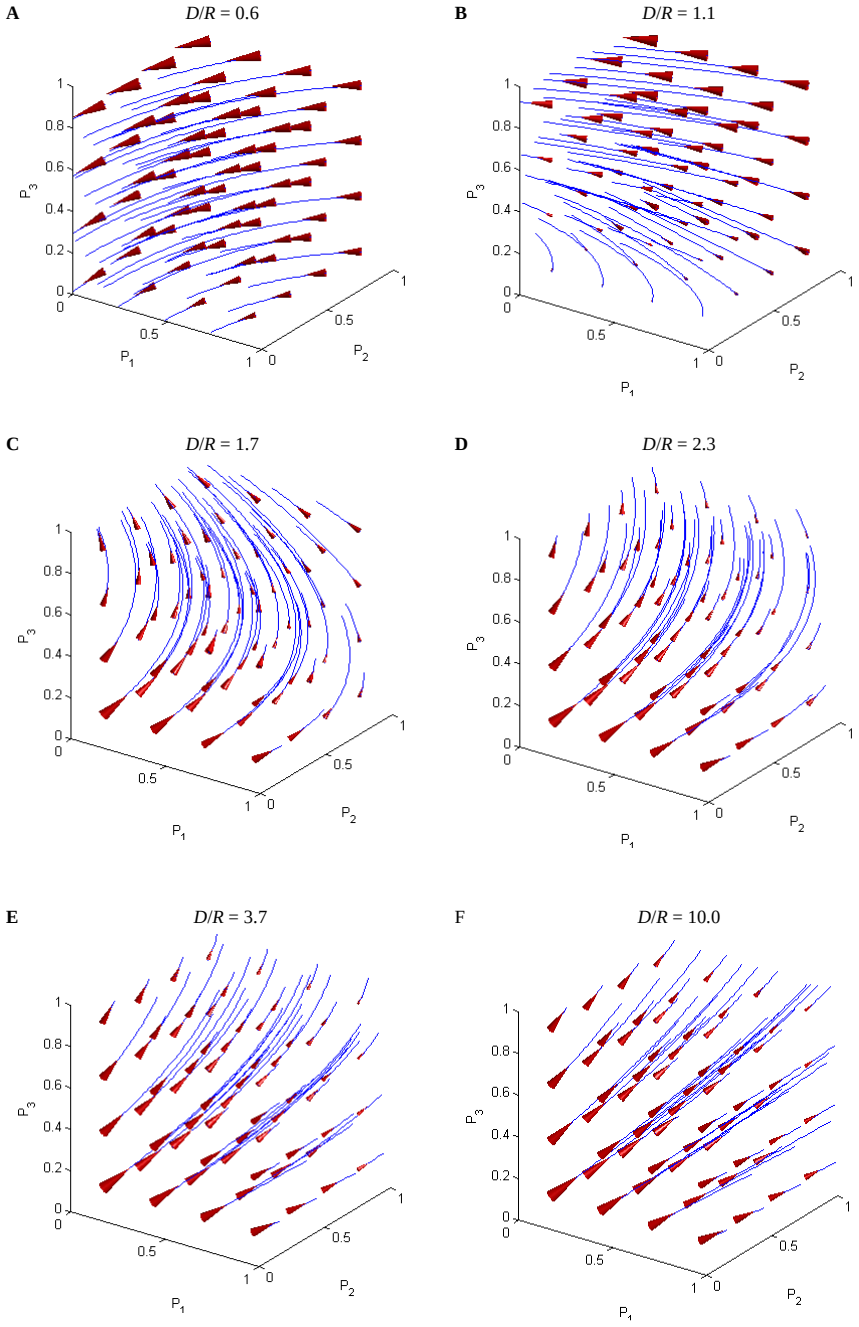

Figure 3.5. Cone arrows show the net direction of movement when n = 3 and *D*/*R* ratio changes from 0.6 to 10.0. (A) *D*/*R* is small. In this case, all agents ignore the intruders. (B, C, D) *D*/*R* is moderately large; the equilibrium is such that Agent 1 ignores (p1 = 0) but Agent 2 and 3 eliminate. (E, F) When *D*/*R* is large enough, Agent 1 starts making elimination effort.

Q-learning is a primitive form of learning (76) in which utility values (*Q* values) are learned for state-action pairs, absent a model of the environment. It provides a simple

means for agents to learn how to act optimally in an unknown environment. At each step, a *Q*-learning agent uses its new experience to improve its long-term reward estimate by combining new information with prior experience.

Each *Q*-learning strategy is determined by the value function, *Q*, which estimates long-term discounted rewards for each action. General scheme of *Q*-learning algorithm used by each agent is as follows:

(1) Observe the current state (in this study, there is only one).

(2) Choose and execute an action based on the *Q*-values from a set of available actions, Act_i (available actions are Act_1 = eliminate and Act_2 = ignore). The agent selects its action according to a probability given by the Boltzmann distribution:

$$p(Act_i) = \frac{e^{Q(Act_i)/T}}{\sum_{Act_k \in Actions} e^{Q(Act_k)/T}} \tag{3.4}$$

where *T*, called "temperature," adjusts the randomness of decisions.

(3) Observe the new state (for this study, this step is not necessary because there is only one state) and receive an immediate reward.

(4) Adjust *Q* value based on the action taken, *a*, using equation (3.5):

$$\begin{aligned} Q(a) &\leftarrow (1-\alpha)Q(a) + \alpha(reward + \beta V) \\ V &= \max_b Q(b) \end{aligned} \tag{3.5}$$

where α is the learning rate ($0 \le \alpha < 1$) and β is the discounting factor ($0 \le \beta < 1$). Here we specified $\alpha = 0.1$ and $\beta = 0.5$. *V* is known as the value of the game and is equal to the maximum *Q* value.

The Agent-Based Model

In mammalian livers, an absorbed toxin can be cleared by any of a sequence of hepatocytes. In order to gain insight into that process, we modeled the liver as consisting of many, parallel sets of toxin eliminating agents arranged in sequence from PV to CV in Figure 3.1. Each agent used a *Q*-learning algorithm to decide its clearance strategy. An agents' task was to minimize the extrahepatic damage to the organism of which they are part. We specified that agents become aware of extrahepatic tissue damage via alarm signals that are quickly released into blood by the damaged tissue. Doing so was based on the fact that hepatocytes, like immune cells, express toll-like receptors (80). They enable cells to detect chemical alarm signals generated by damaged tissues.

Physiologically-Based Model

In addition to the agent-based model, we used a traditional physiologically-based modeling approach to study the effects of hepatic zonation on toxicity exposure to the whole body (Figure 3.6). For simplicity, the liver is represented as having two zones: periportal (zone 1) and perivenous (zone 2). Compounds in the liver are assumed to stochastically take one of the following four paths with probability p_i. Path 1: neither of the two zones encounters the compound. Path 2: only zone 1 encounters the compound. Path 3: both zones encounter the compound. Path 4: only zone 2 encounters the compound.

There is one set of differential equations for each path (Table 3.1). At each time step, one of the four sets is chosen according to the probability associated with the corresponding path. In this model, zone 1 and zone 2 eliminate compounds independent of each other. We specify that the mechanisms of xenobiotic elimination in the liver

(including uptake transport, biliary efflux and metabolism) follow saturable Michaelis-Menten kinetics. Intrinsic clearance of each zone is specified to be $CL_i = V_{max,i}/Km$, where V_{max} is the maximum elimination (metabolic + transport) rate and Km is the Michaelis-Menten constant. V_{max} is assumed to be affected by the level of metabolic enzymes and transporters expressed by cells. As a result each zone has its own V_{max}. The two zones are specified to have equal *Km* values.

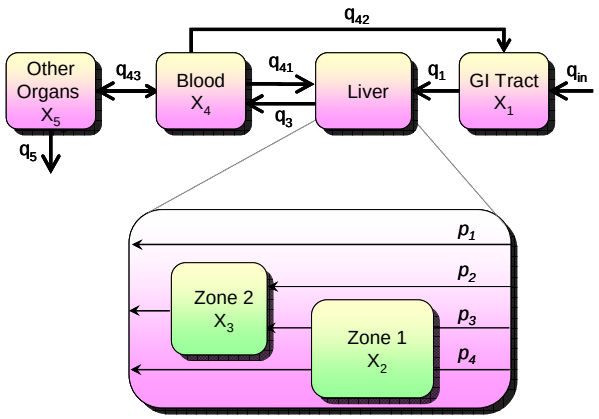

Figure 3.6. A physiologically-based model to analyze the effects of hepatic zonation on toxicity exposure to the whole body. Boxes are reservoir compartments. Arrows show xenobiotic flow directions. q's are xenobiotic flow rate constants. X_i's are xenobiotic concentration in corresponding reservoir compartment shown. p_i is the probability that xenobiotic passes through path i.

Further, we represent the cost to the organism, J, as being proportional to $(CL_1)^2 + (CL_2)^2 + (c\text{AUBC})^2$, where CL_1 and CL_2 are intrinsic periportal clearance and intrinsic perivenous clearance, respectively; AUBC is the area under the blood concentration curve (exposure) and *c* adjusts the relative cost contribution of a fixed dose based on xenobiotic toxicity.

Table 3.1. Equations of the model for each of the four paths shown in Figure 3.6.

Path 1	$V_1 \frac{dx_1}{dt} = -q_1 x_1 + q_{42} x_4 + q_{in}$ $V_4 \frac{dx_4}{dt} = q_1 x_1 - (q_{42} + q_{43}) x_4$ $V_5 \frac{dx_5}{dt} = q_{43} x_4 - q_5 x_5$
Path 2	$V_1 \frac{dx_1}{dt} = -q_1 x_1 + q_{42} x_4 + q_{in}$ $V_2 \frac{dx_2}{dt} = q_1 x_1 - q_3 x_2 - \frac{V\max_1 x_2}{Km_1 + x_2} + q_{41} x_4$ $V_4 \frac{dx_4}{dt} = q_3 x_2 - (q_{41} + q_{42} + q_{43}) x_4$ $V_5 \frac{dx_5}{dt} = q_{43} x_4 - q_5 x_5$
Path 3	$V_1 \frac{dx_1}{dt} = -q_1 x_1 + q_{42} x_4 + q_{in}$ $V_2 \frac{dx_2}{dt} = q_1 x_1 - q_2 x_2 - \frac{V\max_1 x_2}{Km_1 + x_2} + q_{41} x_4$ $V_3 \frac{dx_3}{dt} = q_3 x_2 - q_3 x_3 - \frac{V\max_2 x_3}{Km_2 + x_3}$ $V_4 \frac{dx_4}{dt} = q_3 x_3 - (q_{41} + q_{42} + q_{43}) x_4$ $V_5 \frac{dx_5}{dt} = q_{43} x_4 - q_5 x_5$
Path 4	$V_1 \frac{dx_1}{dt} = -q_1 x_1 + q_{42} x_4 + q_{in}$ $V_3 \frac{dx_3}{dt} = q_1 x_1 - q_3 x_3 - \frac{V\max_2 x_3}{Km_2 + x_3} + q_{41} x_4$ $V_4 \frac{dx_4}{dt} = q_3 x_3 - (q_{41} + q_{42} + q_{43}) x_4$ $V_5 \frac{dx_5}{dt} = q_{43} x_4 - q_5 x_5$

Model Parameters

Volumes of distribution: The apparent volume of distribution differs from compound to compound and from organ to organ. We assume that the model compartments are well-stirred and substrates instantly distribute in the entire tissue volume. The volumes of organs are reported for a 250-g rat in Table 3.2.

V_1 = Volume of G.I Tract (V_{gi}) = $V_{stomach} + V_{small\ intestine} + V_{spleen}$

V_2 = Volume of zone 1 (V_{z1}) = (3/4) V_{Liver}

V_3 = Volume of zone 2 (V_{z2}) = (1/4) V_{Liver}

V_4 = Volume of blood (Vb)= $V_{arterial\ blood} + V_{venous\ blood}$

V_5 = Volume of other organs = $V(muscle + skin + adipose + heart + kidney)$

If the liver lobule is roughly approximated as a cylinder (radius r and height h) with two zones, then zone 2 (perivenous) could be thought of as a smaller cylinder with radius r/2. The rest of the volume would represent zone 1 (periportal). The ratio of the two volumes can be calculated as follows: Volume of zone 1 = $V_{z1} = 2h\pi r^2 - h\pi r^2$; Volume of zone 2 = $V_{z2} = h\pi r^2$; $V_{z1}/V_{total} = 3\pi r^2/4\pi r^2 = 3/4$; and $V_{z2}/V_{total} = \pi r^2/4\pi r^2 = 1/4$.

Table 3.2. Physiological parameters of tissues in a 250-g rat (81).

Tissue	V (ml)	Q (ml/min)
G.I.Tract	13.1	9.8
Liver	10.3	11.8
Blood	16.9	43
Other	176.9	n/a

Rate constants:

q_1= blood flow of G.I. Tract (Q_{gi})

q_3=hepatic blood flow (Q_L)

q_{41}=hepatic arterial flow ($Q_L – Q_{gi}$)

q_{42}=G.I. Tract blood flow (Q_{gi})

q_{43}=effective flow of substrates from blood to other organs that we assume is generally less than the sum total blood flow of the organs.

q_5=clearance rate of xenobiotics from blood by other organs that is primarily done by kidney (renal clearance).

The rate constants are listed in Table 3.3.

Table 3.3. Rate constant values.

Parameter	Value
q_1	9.8 (=Q_{gi})
q_3	11.8 (=Q_L)
q_{41}	2.0 (=Q_L–Q_{gi})
q_{42}	9.8 (=Q_{gi})
q_{43}	1
q_5	0.5

3.1.4 Results

Figure 3.7 shows the emergent strategies obtained for different values of *D*/*R*. When *D*/*R* is small, agents expend little effort to eliminate intruders. As *D*/*R* increases, downstream agents expend more elimination effort than upstream agents. When *D*/*R* is large, upstream agents begin cooperating and contribute to the elimination process, until all agents are expending maximum effort. Although the downstream agents always expend an equal or greater effort than do upstream agents, it does not mean that downstream agents actually eliminate more intruders. For example, when *D*/*R* = 5.0, upstream agents eliminate more intruders than do downstream agents.

Figure 3.8 shows the results from the physiologically based model. Figure 3.8A shows a typical 3D surface of the cost function, J, when c = 0.1. At that toxicity value, the minimum cost (J_{min}) occurs when CL_1 = 0.35 and CL_2 = 0.45 (i.e. zone 2 expends more clearance effort than zone 1). What happens if toxicity is altered? Figure 3.8B shows how J_{min} changes if toxicity varies from 0.05 to 5. It depicts as toxicity increases,

both CL_1 and CL_2 increase however CL_2 is greater or equal to CL_1 at all toxicity levels. That observation is consistent with the game theoretic and multi-agent models results.

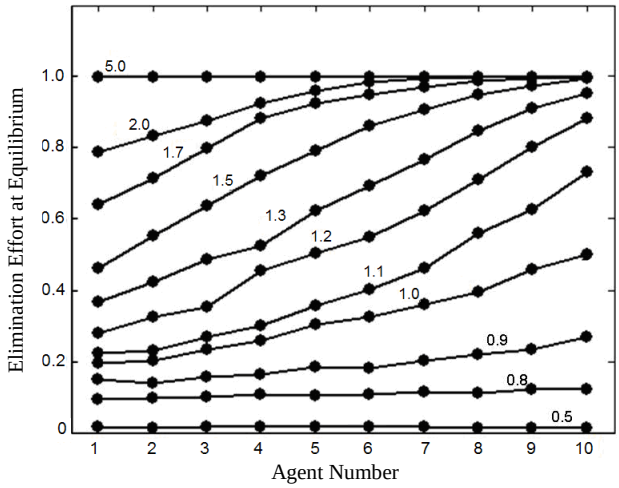

Figure 3.7. Equilibrium of the n-player game changes with *D*. Upstream is to the left (small numbers) and downstream is to the right. The average elimination strategies of 10 agents are shown for different ratios of *D*/*R* (shown on each curve) after 10,000 simulation steps ($maxE_i = 0.05$). When *D*/*R* is small, upstream and downstream agents expend little elimination effort. As *D*/*R* ratio increases, downstream agents expend more elimination effort than upstream agents. When *D*/*R* is large, upstream agents start to cooperate and contribute to the elimination process.

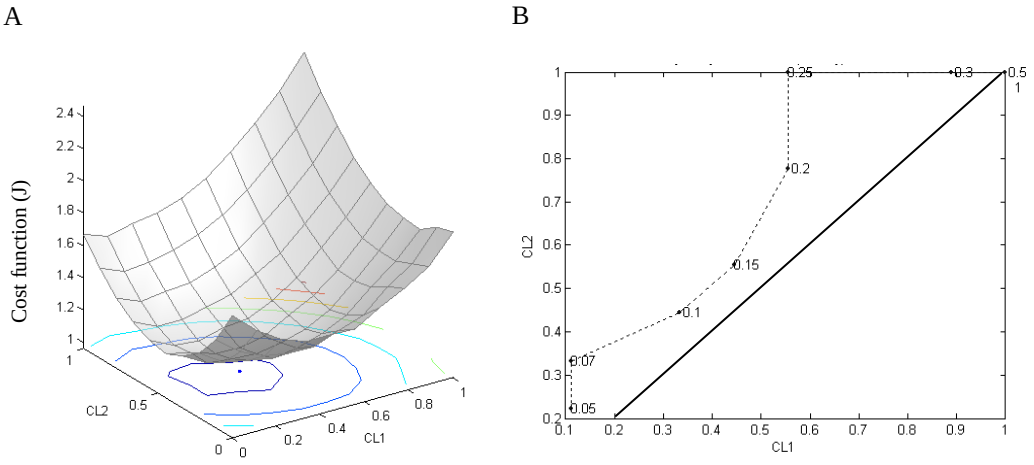

Figure 3.8. Effects of toxicity change in the physiologically-based model. (A) 3D plot of cost function, J, versus CL_1 and CL_2, when $c = 0.1$. (B) Dotted curve: trajectory of J_{min} as *c* changes from 0.05 to 0.5. For each point on the trajectory the value of corresponding c is reported. Solid line: the unity line.

3.1.5 Validation

Liver cells exhibit a similar behavior: they express heterogeneous, location-dependent enzyme and transporter activities to detoxify compounds. Downstream cells commit most to toxin elimination; in other words, xenobiotic metabolism is preferentially located downstream, in the perivenous region (70).

3.1.6 Discussion

We presented a simple, agent-based model of a generalized, hepatic xenobiotic clearance process. The model consists of a group of agents that, similar to hepatocytes, cooperate to protect a commonwealth against toxic intruders. The agents do not have a priori information about either the environment or other agents (e.g., the number of other agents, actions available to them, costs associated with their actions, etc.). The agents use *Q*-learning, a primitive form of learning, to minimize their long-term discounted costs. Agents are assumed to know the cost of their own actions. We also assume that relatively fast communication mechanisms provide appropriate danger signals to agents, informing them about damage caused. Furthermore, agents are assumed to take and use an optimal policy.

Simulations showed that agents adjust their clearance effort based on the following two factors: the potential damage caused by intruders, and their ranked proximity to the entity being protected. Downstream agents (the ones with less proximity to the commonwealth) generally expend more elimination effort than do upstream agents, depending on the threat.

The emergent, collective behaviors of these agents are similar to those of hepatic

cells in terms of xenobiotic clearance. The model suggests that an underlying mechanism responsible for liver zonation may be similar to the model's simple mechanism. Hepatocytes may possess subsystems (e.g. special proteins, signaling pathways, etc.) that produce phenomena that have properties that are similar to those of the *Q*-learning algorithm.

Hepatocytes exhibit complex behavior. Recent studies show that toxin-induced hepatocyte injury is not a simple passive process regulated by the dose of an inducer compound; rather it is an active process in which active signaling plays a crucial role (84). Hepatocytes change protein expression levels in response to toxic shocks (70) and adjust their sensitivity to signaling molecules (for example see (85) and (86)). Upstream hepatocytes can communicate with downstream hepatocytes via blood borne signals (73) and/or intercellular calcium waves (87, 88). On the other hand, downstream cells can communicate back with upstream cells via bile acids (89). This bidirectional communication creates a complex intercellular feedback system which might contribute to regulation of adaptation (or a primitive form of learning) in the liver.

3.2 Computational Experiments Reveal Plausible Mechanisms for Changing Patterns of Hepatic Zonation of Xenobiotic Clearance and Hepatotoxicity[1]

3.2.1 Introduction

Hepatic zonation is evident periportal (afferent) to perivenous (efferent) attribute gradients within lobules. Zonal differences occur in the clearance of a variety of endogenous compounds and xenobiotics (7, 70). Zonation is also evident for a number of normal hepatic functions absent xenobiotic or toxin exposure. There is also differential sensitivity to the induction of cytochrome P450 isozymes (90). Toxin caused hepatic injury can also exhibit zonal patterns. Such phenomena are most often ascribed to having a multifactorial basis, in which oxygen gradients, other blood-borne signals, and blood flow itself may play prominent roles (73, 91, 92, 93). Recent evidence supports the hypothesis that components of the Wnt-β-catenin pathway may play an important role (94, 95, 96). Braeuning (97) reviews the role of several pathways including Ras-Raf-MAPK (mitogen-activated, protein kinase) and Wnt-β-catenin. However, no concrete, causal, mechanistic theory has yet been offered for how different types of hepatic zonation phenomena emerge following dosing with different compounds. For this study, we focused on zonation patterns of P450 isozymes and the hepatic damage that can develop following treatment of rats with xenobiotics. We used the synthetic method of modeling and simulation (2) to discover, explore, and experimentally

[1] Submitted for review, Journal of Theoretical Biology

challenge concrete mechanisms that show how and why biomimetic zonation patterns emerge and change within an agent-based analogue of a hepatic lobule in response to compound dosing. The *in silico* mechanism may have counterparts in rats.

Christoffels et al. (73) demonstrated the plausibility of a molecular level mechanism for periportal-to-perivenous gradients of gene expression. Expanding upon the zonation ideas offered in (7), they hypothesized that interaction between two or more, different signal gradients is necessary to enable development of periportal-to-perivenous gene expression patterns that mimic those gradients and are stable under different conditions. They provided support for the hypothesis using both an inductive, mathematical model and a transgenic mouse model into which hepatocyte-specific DNA-response units had been integrated. They discuss the formidable issues of constructing such model systems using transgenic mice.

Ohno et al. (120) constructed sophisticated, single-hepatocyte based lobular models that focus on ammonia metabolism with the long-range objective of elucidating how molecular and cellular level properties modify higher-level phenomena. Xenobiotic metabolism and enzyme induction mechanisms were not a focus. They posit that heterogeneous gene expression evolved to optimize energy efficiency. They specify histological structure and zone-specific gene expression of major enzymes, and include the biochemical kinetics of enzymes and transporters. Ierapetritou et al. (121) recently reviewed the models of Ohno et al. along with a variety of additional computational liver models. Several specified features of zonation but not how those features may emerge. All but one of the models reviewed was an inductive mathematical model. The computational modeling and simulation (M&S) approach used herein (Figure 3.9a) and

the resulting models are fundamentally different from those inductive mathematical models and so are not directly comparable. Hunt et al. (2) explain those differences and how the two different M&S approaches complement each other.

The approach used herein (Figure 3.9A) was developed to enable construction of biomimetic mechanisms that are real (not conceptual) and strictly defined, and conditions are less supportive of inductive modeling methods (Figure 3.9B). Even though abstract, the mechanisms and their spatial context are flexible and sufficiently concrete to instantiate mechanistic hypotheses and test their plausibility experimentally.

Following cycles of model construction, evaluation and selection, and refinement, we arrived at a discrete event, discrete time system that maps to a cross-section through a hepatic lobule having periportal-to-perivenous flow and a connection to extrahepatic tissue. The hepatic lobule component is comprised of 460 identical, quasi-autonomous functional units called sinusoidal segments (SSs). Each SS maps to a small portion of a sinusoid (Figure 3.10). During a simulation cycle, each SS has an opportunity to clear a detected compound administered at the start of the simulation. A compound that exits the lobule enters an extrahepatic tissue space and causes release of a response signal, proportional to potency. SSs detect and respond to response signals. The probability of a SS clearing a compound depends on the current clearance effort (defined in Methods) of that SS. The SS also observes the local level of an endogenous gradient and reacts to any response signals. Alternatively, the response signal can arise from the SS or the compound itself could fill the role of response signal. Each SS adapts to the new information with the objective of improving efficiency: it can reduce response signals by

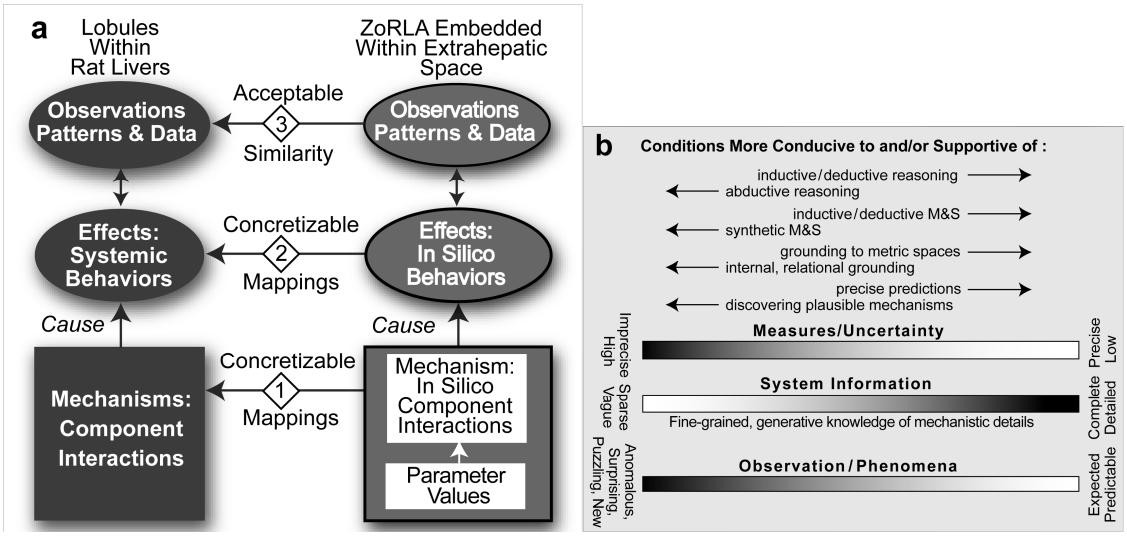

Figure 3.9 Features of the synthetic method of modeling and simulation. (a) Shown are relationships between observations made on rat liver lobules following xenobiotic treatments and the *in silico*, zonally responsive LOBULAR analogues (ZoRLA) in Figure 3.10 following simulated treatments with xenobiotics. Left: the referent systems are experimental observations made on rat livers following various treatments with xenobiotics. During experiments, lobular components interact with administered xenobiotics causing the compound to be cleared (metabolized or excreted in bile). The compound may also generate tissue responses, both intra- and extrahepatic. The consequences of localized mechanistic events cause systemic effects. Systemic behaviors at all levels are reflected in the recorded data. Right: Abstract, software components are designed, coded, verified, assembled, and connected. The product of the process is a ZoRLA (as in Figure 3.10) within an EXTRAHEPATIC space. Concretizable mappings exist between ZoRLA components and how they plug together, and lobular physiological and functional detail at the level of a sinusoid, as illustrated in Figure 3.10. Execution gives rise to a working analogue; measures of events provide results. Dynamics during execution (mappings 2) are intended to represent abstractly plausible corresponding dynamics (believed to occur) within the rat during an experiment. Measures of dynamics—patterns of zonation, in this case—provide data that may or may not mimic wet-lab counterparts. Achieving measurable similarities makes mappings 3 quantitative. (b) Conditions supportive of both the synthetic method of modeling and simulation (M&S) along with the familiar inductive method of M&S are sketched. Obviously, toxicologists and pharmacologists would like explanatory knowledge about hepatic phenomena and zonation to be rich and detailed, and for uncertainties to be limited. Such conditions (toward the far right side), which are common for non-biological, engineered systems, favor developing inductive models that are increasingly precise and predictive. However, as discussed in (2) absent detailed knowledge of the causes of hepatic zonation, we are on the left side, where frequent abduction is needed and synthetic M&S methods can be most useful.

increasing clearance effort in subsequent rounds or reducing clearance effort when no response signals are detected. To do so it uses the local gradient and response signal information to update its clearance effort using a learning algorithm.

Upon compound exposure, the lobular component developed a variety of periportal-to-perivenous (P-to-P) clearance effort gradients. Several gradient patterns were strikingly similar to those reported in the literature for P450 isozymes following xenobiotic dosing (for convenience, examples are provided in Appendix C Figure C.10). Zonal patterns of clearance effort and SS damage changed depending on compound dose and potency. We called the system a zonally responsive lobular analogue (ZoRLA). A ZoRLA was used to achieve a degree of quantitative validation against data on hepatic zonation of CYP1A2 mRNA expression caused by three different doses of TCDD (2,3,7,8-tetracholorodibenzo-p-dioxone).

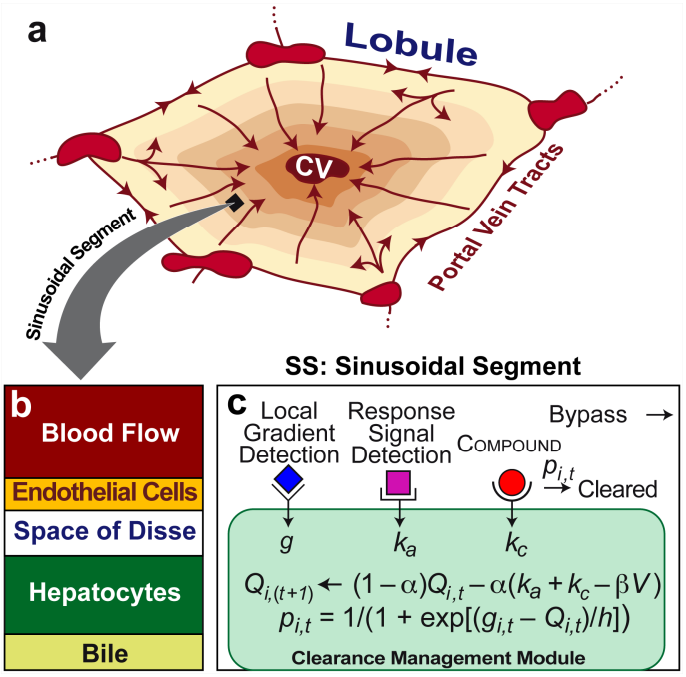

Figure 3.10 Referent and analogue. (a) An illustration of a cross-section through a hepatic lobule showing sinusoidal flow paths from portal vein tracts to the central vein (CV) and a functional periportal-to-perivenous zonation pattern. (b) A ZoRLA is an arrangement of Sinusoidal Segments (SSs). A SS is a quasi-autonomous agent. As illustrated here, it maps to small portion of a lobule that includes portions of the sinusoidal blood flow and space of Disse, along with one or more endothelial cells and hepatocytes. (c) A SS can detect and act on co-located mobile objects and signals. Three object/signal types are used: those that comprise the local gradient, response signals (generated by a compound's extrahepatic response), and a COMPOUND. The local value of the gradient is g. k_a is the COMPOUND'S potency. To reach the CV, a COMPOUND must pass through a sequence of SS and escape being cleared. If a COMPOUND is undetected by the CLEARANCE Management Module, it exits (bypass). Upon detection, the COMPOUND is cleared with probability $p_{i,t}$ (simulation cycle t). If not cleared, it exits. The value of Q used by SS_i during simulation cycle t is an estimate of its long-term, discounted cost of continuing to use its current clearance effort, p_t. The current clearance strategy of SS_i is $p_{i,t}$; it is updated as specified in each simulation cycle using that location's g value along with $Q_{i,t}$. $Q_{i,(t+1)}$ is the value to be used during the next simulation cycle. h, k_c, α, β, and V are defined in the text.

3.2.2 Methods

To distinguish clearly *in silico* components and processes from corresponding rat counterparts, we use SMALL CAPS when referring to the former. Parameter names are italicized.

The attributes targeted

The first consideration in any modeling effort is to determine why the models are being created, and identify situations in which they will be used (3). For this project, we sought spatially organized, biomimetic mechanisms that could produce phenomena similar to attributes listed in Table 3.4. Unlisted hepatic phenomena are, for the time being, outside the scope of this project. However, a requirement was that the models and their components be sufficiently flexible so that during a future extension of this project, or when other investigators use the models, they could be easily modified to account for an expanding list of attributes. Clear statements about use and targeted attributes facilitate selecting specifications. Specifications, paired with attributes, are provided in Table 3.4. Clear specifications guide model design and development and help one avoid potentially unproductive tangents. Each attribute achieved provides a degree of validation.

To achieve targeted attributes, we followed an iterative construction and refinement protocol similar to that detailed most recently in (2) and (98). The objective was to discover CELL level mechanisms that would make ZoRLA patterns of clearance effort (defined below) following COMPOUND dosing, increasingly biomimetic, with the long-term goal of achieving ZoRLAs that exhibit all of the diverse set of attributes listed

in Table 3.4, and do so while adhering to a strong parsimony guideline. For this project, we focused on the first ten attributes. For each attribute targeted, we provided a ZoRLA specification that would be commensurate with the other specifications in enabling eventual achievement of all Table 3.4 attributes. The iterative refinement protocol cycles through the following eight steps.

1) Choose an initial, small subset of attributes to target, e.g., 1–3 in Table 3.4.

2) Select a granularity level that will enable comparing measures of simulated and targeted attributes. We selected the level illustrated in Figure 3.10 and Figure 3.11.

3) For each attribute targeted, specify a desired level of phenomenal similarity (e.g., within ± 25% or exhibit the same P-to-P gradient trend). Approach in stages: begin with relaxed similarity measures.

4) Posit coarse-grained, discrete mechanisms that may generate analogous phenomena while requiring as few components as is reasonable.

5) Create logic to be used by each component. Instantiate components and mechanisms. Update specifications.

6) Conduct many, simulation experiments. Measure a variety of phenomena to establish *in silico* to wet-lab similarity and lack thereof.

7) Tune (parameterize) to achieve analogue similarity specified at step 3. When the effort fails, return to step 4. When successful, return to step 3 and increase the stringency of the similarity measure.

8) Add one or more new attributes until the current analogue is falsified. Return to

step 2. Strive to achieve the expanded attribute list with as little component reengineering as possible.

The synthetic method of modeling and simulation

The method used herein is a relatively new experimental approach to discover and challenge plausible, biomimetic mechanisms. It acts on the dictum of the great physicist Richard Feynman "what I cannot create, I do not understand." To gain insight into plausible generative mechanisms that may be responsible for biological phenomena of interest, such as hepatic zonation patterns, when uncertainty is large and detailed data are limited, the method involves building extant—real, actually existing, and observable—working mechanisms that exhibit some of those same phenomena, and thus may be biomimetic. The approach is based on the scientific principle (99) illustrated in Figure 3.9A. When two systems, hepatic lobules of laboratory rats and a model—a ZoRLA—are composed of interacting components for which similarities can be established at some level of abstraction (mappings 1 in Figure 3.9A), and the two systems exhibit several measurable, phenotypic attributes (mappings 3), for which some degree of similarity exists, then there may also be similarities in the generative mechanisms responsible for those attributes (mappings 2). We cannot yet build hierarchical mechanisms out of biochemicals. However, as described herein and recently reviewed (2, 100, 101, 102), we can build extant biomimetic mechanisms using object-oriented software tools. In so doing, as Figure 3.9 illustrates, we are *not* following the traditional, inductive approach of modeling the data. Nor are we describing mathematically the behaviors of a hypothetical, conceptual mechanism. Rather, we explore the space of relatively simple, plausible, biomimetic, mechanisms, for *concrete* instances, which upon

instantiation, verification, and simulation exhibit phenomena, which when measured produce data that are essentially indistinguishable from comparable wet-lab data. Because the MECHANISMS are intended to be analogous to referent biological mechanisms, we refer to the *in silico* system as an *analogue*. To emphasize their concrete, constructive nature (a whole assembled from stand alone parts), we call them *synthetic* analogues. The approach provides an important, scientific and experimental means to explore and test *in silico* plausible, biomimetic mechanistic hypotheses, when it would be too difficult, too expensive, infeasible, or unethical to do so in animals or *in vitro*.

ZoRLA design considerations

In related work, we built *in silico* liver analogues to challenge mechanistic hypotheses and gain improved insight into plausible micro-mechanistic details of xenobiotic clearance (52, 9, 103), hepatic drug interactions (98), diseased-caused differences in spatiotemporal micro-mechanisms influencing hepatic drug disposition (104), and heterogeneities in intralobular enzyme induction (10). At the start of this project, we drew on these methods and their validated components to instantiate and experiment on lobular analogues targeting subsets of attributes 1–10 in Table 3.4.

All ZoRLA system components and processes are discrete. Time advances discretely by simulation cycles and steps within cycles. Each simulation cycle, every component updates its state based on changes since the last update opportunity. Each SS is a software agent. An agent is a quasi-autonomous software object that can schedule its own events. SSs can be connected together in different ways. Within a simulation cycle, SS components interact with mobile COMPOUNDS percolating through a sequence of SS. The process maps to absorbed xenobiotics percolating through

sinusoids and interacting (or not) with spaces and cells as described in (9, 104). Most events are stochastic. During a simulation cycle, when an event opportunity for a given component occurs, there is a parameter-specified probability for each, prespecified outcome.

A specific ZoRLA instantiates a mechanistic hypothesis (2, 101): this component arrangement and operating principles will, upon execution, produce phenomena similar to targeted attributes. Execution followed by comparison of results to referent data tests the hypothesis. The process is directly analogous to constructing and then experimenting on a wet-lab model to test hypotheses. We adhered to a parsimony guideline and strove to keep ZoRLA components as simple as feasible, while achieving the attributes in Table 3.4.

Inductive models are typically grounded to metric spaces. ZoRLAs use relational grounding. Grounding is defined as the units, dimensions, and/or objects to which a variable or model constituent refers. Each ZoRLA component is grounded to a subset of the other components. The merits and limitations of relational grounding are discussed in (2). Using relational grounding means that a separate mapping (transformation of data) model is needed to quantitatively relate measures of ZoRLA phenomena, such as clearance effort, to corresponding wet-lab phenomena, such as a measure of P450 isozyme levels within a tissue sample or isolated cells.

SS intrinsic clearance

The primary phenomena in Table 3.4 are intralobular changes in compound clearance and the intracellular levels of P450 isozymes (or their mRNA) responsible for compound

clearance. We needed a SS counterpart to intrinsic clearance. We can see from Figure 3.10b that some of the compound entering a SS can exit without entering hepatocytes. A COMPOUND that is detected by a SS maps to compound entering hepatocytes. A COMPOUND that is undetected exits that SS; we say it *bypassed* that SS. We specified a bypass probability (*pBypass*) for each SS that could be adjusted for each COMPOUND; with one exception, for the ZoRLA experiments described herein, we arbitrarily set *pBypass* = 0.8. A COMPOUND maps to a small fraction of the xenobiotic dose used in one of the experiments cited in Table 3.4. A COMPOUND that does not bypass is detected; that maps to compound entering cells. We specified that if detected, the COMPOUND would be cleared with probability *p*; that maps to the intrinsic clearance by hepatocytes within a corresponding sinusoidal segment for the time interval to which a simulation cycle maps. We define $p_{i,t}$ to be the clearance effort of SS_i during simulation cycle *t*. Within hepatocytes, an increase in intrinsic clearance typically correlates with an increase in P450 isozymes. For this report, we limit attention to those cases. That increase comes at a cost to the cell. In SS, the cost of a COMPOUND clearance event is k_c. Given the many functions carried out by hepatocytes, we assumed that hepatocytes have an evolution imposed, genetic mandate to avoid unnecessary costs.

Signals and methods needed and used by SS

Gebhardt argued that two classes of signals are necessary and essential for creating zonation (7). An early task was to discover and demonstrate a discretized mechanism that would enable a SS to adjust its clearance effort based on local levels of two signals. Further, within a ZoRLA, when needed, these signals could correspond to amounts of two different mobile objects, RESPONSE (R-) and B-SIGNALS. We specified that R-SIGNALS be

connected to COMPOUND, dose, and potency. We identified three options: 1) a COMPOUND or its METABOLITE is the R-SIGNAL; 2) a COMPOUND or its METABOLITE initiates a response within a SS and that response (or an object generated by that response) functions as a R-SIGNAL that is also detectable subsequently by other SS; and 3) a COMPOUND or its METABOLITE that escapes HEPATIC clearance initiates an EXTRAHEPATIC response, and that response (or an object generated by that response) is an R-SIGNAL that is SS detectable. We could achieve our objective if any of the three options worked. Early exploratory simulations indicated that the third option would be the simplest to implement. Consequently, we focused on that option. Option three may map to a xenobiotic that causes extrahepatic tissue toxicity. Our attention is on interactions between a tissue and an active xenobiotic (or its active metabolite) that cause observable toxic or pharmacological responses. We conjectured that some of those interactions would cause release of factors that are carried by blood to the liver where they function similar to alarm signals (105). Such factors could be detected by hepatocyte Toll-like (80) or other receptors. The algorithm in Figure 3.10c can be modified to give the same clearance effort change for the second option, as well. In early SS (see description of preZoRLA1 in Appendix C), mobile objects served as R-SIGNALS. In the final, 2D ZoRLA discussed below, to reduce the duration of each simulation cycle and thus simplify ZoRLA function, we specified that all SS receive R-SIGNALS at the end of each simulation cycle.

For simplicity, we specified that B-SIGNALS, which are unrelated to xenobiotic, map to something external to the liver and that they are carried to the liver by blood. It could, for example, map to O_2 levels. Again, for simplicity, we specified that its input rate be

constant. We gave each SS the ability to respond to B-SIGNAL objects exactly as they respond to a COMPOUND. There was a bypass probability and a non-zero probability of being cleared when detected. For a single sequence of SS, the detected level of B-SIGNALS would provide information about relative location in the sequence. In early SS (see description of preZoRLA1 in Appendix C), mobile objects, entering the ZoRLA on the periportal side and flowing in the P-to-P direction, functioned as B-SIGNALS. In the final, 2D ZoRLA discussed below, to reduce the duration of each simulation cycle and thus simplify ZoRLA function, we specified that a fixed P-to-P B-SIGNAL gradient exists so that the B-SIGNAL for each SS_i is constant for the duration of the simulation experiment. We used an equation (defined below) to specify the value, g_i, for each SS_i.

Enabling SS to adapt their clearance effort in response to dosing with different COMPOUNDS

Christoffels et al. (73) and others have posited that zonation is linked to the adaptive ability of the liver and hepatocytes specifically. The SS described above are very simple agents. Each SS has three actions: detect B-SIGNALS, detect R-SIGNALS, and clear COMPOUNDS. We conjectured that if a cost were associated to each action, then when given a simple learning algorithm, a SS could use it to avoid increasing costs, and that would result in location dependent clearance efforts. Consequently, a mandate of each SS was to alter clearance effort to avoid rising costs. Because B-SIGNALS are independent of COMPOUNDS and, once the gradient stabilizes, they are constant for each SS, we elected to ignore them, unless we failed to achieve targeted attributes. We specified that the cost of clearing one COMPOUND is k_c. We specified that the cost of detecting one R-SIGNAL is k_a. We conjectured that if a xenobiotic disrupts normal tissue

function or causes tissue damage, the consequences of such an event could cause hepatocytes to alter their behaviors, and so doing would come at a cost to the hepatocytes. The relationship is more clear when it is the hepatocyte itself that is disrupted or damaged. Increasing p increases costs. Costs are lowest when a SS chooses $p = 0$. However, if the COMPOUND is a TOXIN, then with $p = 0$, all COMPOUNDS would exit the LOBULE causing extrahepatic responses. Subsequent R-SIGNAL detection by all SS would increase costs. If there is a direct relation between TOXIN potency and the number of R-SIGNALS detected (or their value), then it is clear that P-to-P patterns of clearance effort would change as COMPOUND dose and potency are changed. We explored those changes. Based on recent observations about simulated hepatocyte learning in Section 3.1 (68, 69), we anticipated that, given a simple learning mechanism (an algorithm in this case), a SS at the end of simulation cycle $t - 1$ could find a new p_t that would be expected to lower costs during subsequent simulation cycles.

The SS mechanism in Figure 3.10c provides the above capabilities and is simple enough so that there may be one or more yet to be identified hepatocyte counterparts. The figure shows an SS with a quasi-autonomous subsystem, the CLEARANCE Management Module. It maps to all hepatocyte resources and subsystems associated with xenobiotic clearance. The CLEARANCE Management Module has one mandate: adjust p_t up or down or keep it the same to lower future costs. A function in form of a Boltzmann distribution is provided that adjusts $p_{i,t}$.

$$p_{i,t} = 1/(1 + \exp[(g_i - Q_{i,t})/h]) \quad (3.6)$$

$Q_{i,t}$ is based on the well-known Q-learning algorithm (76). The value of Q used by SS_i

during simulation cycle t is an estimate of its long-term, discounted cost of continuing to use its current clearance effort, p_t. At the start of each simulation cycle, p_t is calculated using the Q_t value carried-forward from the previous ($t - 1$) simulation cycle (Eq. (3.6)). At the end of the current simulation cycle, $Q_{i,t}$ is updated using Eq. (3.7). Q_{t+1} is then carried forward for use in the next simulation cycle. h is a constant. The parameter h acts as a gain: it amplifies and/or dampens the difference: $g - Q$. A large value of h eliminates zonation completely ($p_{i,t}$ is constant at 0.5.); a small value magnifies it. When h is very small, $p_{i,t}$ essentially takes one of two values (switch-like behavior).

$$Q_{i,t+1} = (1-\alpha)Q_{i,t} - \alpha(k_a + k_c - \beta V) \tag{3.7}$$

where $V = \min(Q_{i,t}, g_i)$ (see Section C.2 in Appendix C for derivation details). β is known in the Q-learning literature as the discount factor; it determines the current importance of future costs. For the SS in Figure 3.10, as $\beta \to 0$, the difference between PERIPORTAL and PERIVENOUS $p_{i,t}$ values increases. k_a = potency. α is a constant in [0, 1] that controls the CLEARANCE Management Module's learning rate. For the results presented, β = 0.5, k_c = 1, α = 0.1 and h = 0.01, 5, 10, or 70, depending on the experiment. As specified above, g_i is the value of the P-to-P B-SIGNAL gradient used by SS_i. We used a simple equation,

$$g_i = (X_i^2 + Y_i^2) + \eta \tag{3.8}$$

where X_i and Y_i are coordinates of agent SS_i with respect to the center of the grid space, and η is uniform random noise in the interval [0, m/2]; we used m = 31; that is the number of spaces per grid edge.

At the end of each simulation cycle, there are two types of SS: those that 1) did and 2) did not see a COMPOUND. In the first case, the SS performed an action: it either ignored or cleared the COMPOUND. Only SSs of the first type update their *Q* value. If a SS saw one COMPOUND during the simulation cycle, it updated *Q* once. If it saw *n* COMPOUNDS during the simulation cycle, it updated *Q n* times.

When a SS detects a COMPOUND, it is either cleared or not. When a COMPOUND has been cleared, $k_c = 1$, else $k_c = 0$. When an R-SIGNAL is detected, k_a (a measure of potency) takes a value > 0, else $k_a = 0$. The latter still applies in the special case where the COMPOUND is the R-SIGNAL. The term $(k_a + k_c)$ represents the per-COMPOUND cost to each SS for the actions taken during simulation cycle *t*.

SS arrangement and COMPOUND movement

A LOBULE is a system of SSs arranged to mimic a cross-section through a hepatic lobule. The 2D square grid shown in Figure 3.11a maps to a 2D view of a cross section through a hepatic lobule. Located at each grid space is a SS. COMPOUNDS enter from the external PERIPORTAL area, which maps to portal vein tracts. If not cleared, COMPOUNDS move in discrete steps, from the exit of one SS to the entrance of an adjacent SS. Event logic from the COMPOUND'S perspective is diagrammed in Figure 3.11b. COMPOUNDS move stochastically toward the central ZoRLA space, which maps to a lobule's central vein. The process mimics blood flow through sinusoids. Sinusoid interconnections enable some lateral periportal flow. There is no lateral flow in perivenous sinusoids. We implemented and verified a movement algorithm to enable

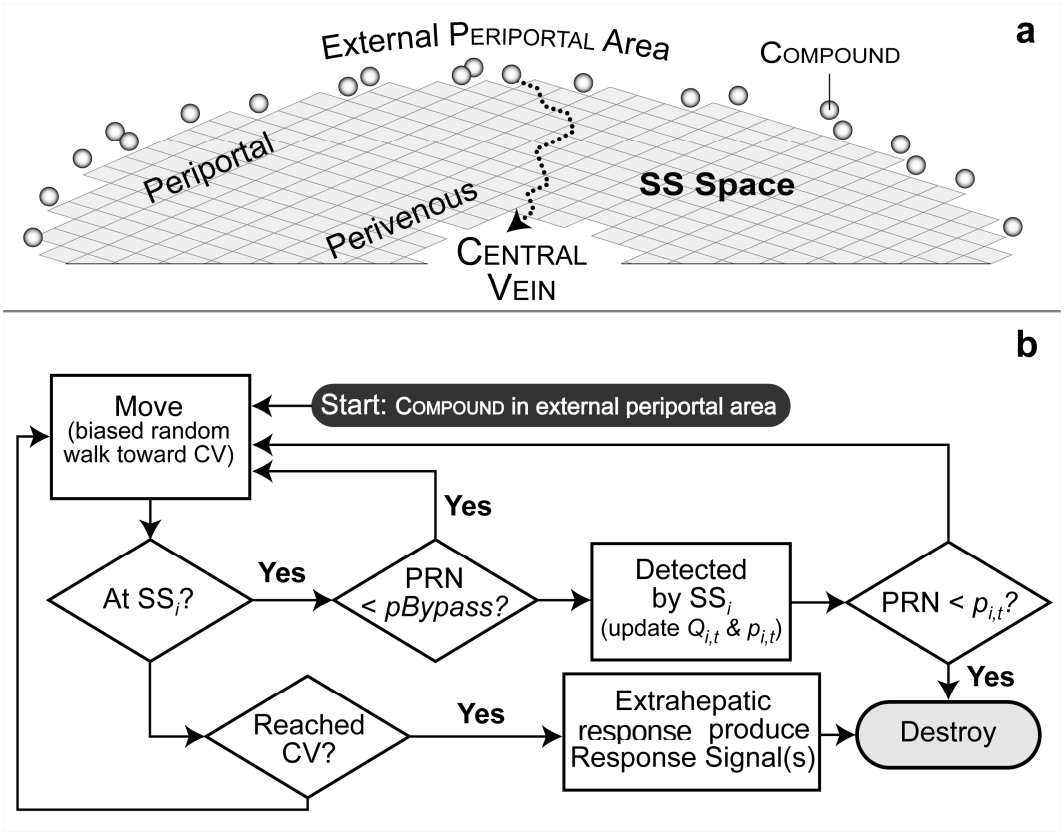

Figure 3.11 Events occurring within simulation cycle t. (a) Illustrated is a portion of a ZoRLA; SS agents (gray squares) are arranged into the illustrated 2D semi-circular grid pattern. The external periportal area maps to portal vein tracts. At the start of each simulation cycle, the compound dose is placed outside. Each compound enters at a randomly assigned SS on the outside edge. During each simulation cycle, compounds move towards the central vein (CV: the central, vacant region; it maps to both central vein and extrahepatic tissues) using the biased random walk described in Methods. The dotted arrow is an example of a compound's path. When a compound (ka > 0) reaches the central vein it is removed, a response occurs, and r-signals are generated in proportion to ka. (b) The flowchart shows event logic from a compound's perspective. During simulation cycle t, the Clearance Management Module detects a co-located compound with probability = 1– pBypass. Upon detection, it is cleared (removed) with probability $p_{i,t}$. At the end of each simulation cycle, each SSi updates its $Q_{i,t}$ (to $Q_{i,(t+1)}$) and clearance strategy (pi,t) based on the number of compounds cleared and the number of r-signals received. PRN: pseudo random number in [0,1].

COMPOUNDS to follow a biomimetic one-grid-space-at-a-time path toward the CENTRAL VEIN. Each simulation cycle, the algorithm below specifies a move direction θ for each COMPOUND. The COMPOUND then uses θ to select, move to, and then enter an adjacent SS. All neighboring grid locations were numbered counter clock wise, starting with 1 at 0° and ending with 9 at 360° (see Figure C.4A). We used a Normally distributed random variable (x) for which the mean ($\mu \in [1,9]$) corresponds to the effective flow direction, θ. For example, $\mu = 4$ corresponds to $\theta = 135°$ and $\mu = 1.5$ to $\theta = 22.5°$. More formally, $x \sim N(\lfloor \theta/45 \rfloor + 1, \sigma^2)$, where $0° \leq \theta < 360°$ is the effective angle of movement (illustrated in Figure C.4A) and σ is the standard deviation of x; σ controls each COMPOUND'S lateral movement.

A COMPOUND remains in place with probability p_{stop}; or it moves stochastically to an adjacent SS with probability $1 - p_{stop}$. The adjacent space, *nextSpace*, is stochastically chosen as: $nextSpace = \lfloor x - 0.5 \rfloor$(modulo 8) +1. Distributions are shown in Figure C.4 for $\theta = 135°$ and three values of σ. The motion converges to Brownian motion for large values of σ (because the x distribution becomes effectively uniform). We used three σ values: σ_1, σ_2 and σ_3, for PERIPORTAL, middle and PERIVENOUS zones, respectively. Because the sinusoid interconnections and therefore the lateral movement is greater periportal, we arbitrarily set $\sigma_1 = 1.2$, $\sigma_2 = 1.0$ and $\sigma_3 = 0.7$ (illustrated in Figure C.4).

As a simulation progresses, we can observe how the P-to-P p_t patterns change and when they stabilize. The hypothesis being tested is that upon sustained COMPOUND dosing, use of the mechanism in Figure 3.10c can produce stable p_t values that exhibit P-to-P patterns. We cannot rule out that some patterns may have no observed biological counterpart. Our

interest herein is in patterns that are biomimetic.

Initial configurations

Given any pre-existing pattern of clearance effort ($p_{t,i}$) and a specific Eq. (3.7) parameterization, the same new pattern will eventually emerge upon sustained exposure to a new COMPOUND. However, the number of simulation cycles required can vary considerably depending on the pre-existing pattern. For simplicity, Q_i values were initialized at the start of all simulations such that Eq. (3.6) produced random values of clearance effort, as follows. $Q_i = g_i + h \cdot ln(1/\text{PRN} - 1)$, where h is an adjustable parameter (the same as h in Eq. (3.6)), g_i is calculated using Eq. (3.8) and PRN is a pseudo-random number between 0 and 1. Thereafter, each SS used Eq. (3.7) to calculate $Q_{i,t+1}$, which was then used in Eq. (3.6) to calculate $p_{i,t+1}$.

3.2.3 Results

Results from two ZoRLA predecessors

Two ZoRLA predecessors are presented in Appendix C. The earlier of the two, preZoRLA1, used a multicomponent positive and negative feedback mechanism within each SS to form and respond to gradients of R- and B-SIGNALS (targeted attribute 3). The mechanism also produced and removed objects called PROTEINS. Clearance effort (p) for each SS was proportional to current number of PROTEINS in that SS. The mechanism was capable of forming a variety of P-to-P B-SIGNAL and clearance effort gradient patterns that were both dose and potency dependent (for examples, see Figure C.5). Parameterizations of preZoRLA1 enabled achieving targeted attributes 1–6. However, inclusion of attribute 7 in the set of currently targeted attributes falsified preZoRLA1.

A second early model, preZoRLA2, comprised a simple linear sequence of 20 of the SS illustrated in Figure 3.10. It maps to a single P-to-P sinusoidal flow path. It too was capable of forming a variety of P-to-P B-SIGNAL and clearance effort gradient patterns that were both dose and potency dependent (for examples, see Figure C.6), and it achieved attribute 7. During preZoRLA2 experiments we tracked the number of COMPOUNDS actually cleared in addition to the current clearance effort of each SS (the values are graphed in Figure C.6). Often, the two patterns were not the same. Consider compounds that are hepatotoxic. Limit attention to those compounds for which toxicity is proportional to hepatocyte exposure. We can reasonably infer that hepatocyte exposure is directly correlated with total amount of compound cleared by that hepatocyte. Given that, the number of COMPOUNDS actually cleared by a preZoRLA2 is directly correlated with HEPATOTOXICITY. COMPOUNDS that were not cleared still caused extrahepatic responses. We also specified that SS HEPATOTOXICITY would ensue when SS lost the ability to detect (and clear) COMPOUND, an event that was set to occur when the number of cleared COMPOUNDS passed an arbitrary threshold (e.g., 100). HEPATOTOXICITY can map to severe hepatotoxicity or even to localized necrosis.

Zonation patterns during 2D ZoRLA experiments depend on COMPOUND potency

To explore emergence of stable zonation patterns, we conducted two sets of experiments using somewhat different ZoRLAs, one using $h = 5$, and the other using $h = 70$. Both used the same prespecified P-to-P g gradient for all five experiments. In both cases, ZoRLAs were dosed at the start of each simulation cycle with 50 COMPOUNDS having the same potency. Except for repeat experiments, potency was changed for different experiments. A simulation cycle continued until all COMPOUNDS exited or

were cleared. Several experiments were conducted for each k_a. In all experiments, at t = 0, random p values were assigned to each SS. Values at early t for repeat experiments (same k_a) were not identical, but they became identical by the end of the experiment.

In the first set of experiments (Figure 3.12) h = 5, and COMPOUND potency spanned a 20-fold range (k_a = 1–20). Experiments terminated at t = 1,000. SSs were assigned to ten equal width zones. Values of p_i, Q_i, and the cumulative number of COMPOUNDS cleared by each SS in a given zone (COMPOUND elimination count) were averaged at t = 1,000. The value of Q used by SS_i during simulation cycle t is an estimate of its long-term, discounted cost of continuing to use its current clearance effort, p_t. The results graphed in Figure 3.12 show that increasing potency caused values of p_{avg} for periportal zones to increase; it also caused the peak COMPOUND elimination count (which maps to measurable hepatotoxicity) to shift from PERIVENOUS to PERIPORTAL. Within each zone, Q_{avg} increased with increasing potency.

In the second set of experiments (Figure 3.13) h = 70, and COMPOUND potency spanned a 10-fold range (k_a = 0–10). Experiments were terminated at t = 500 once stable patterns were achieved. The different, individual SS clearance effort values are shown in Figure 3.13b.

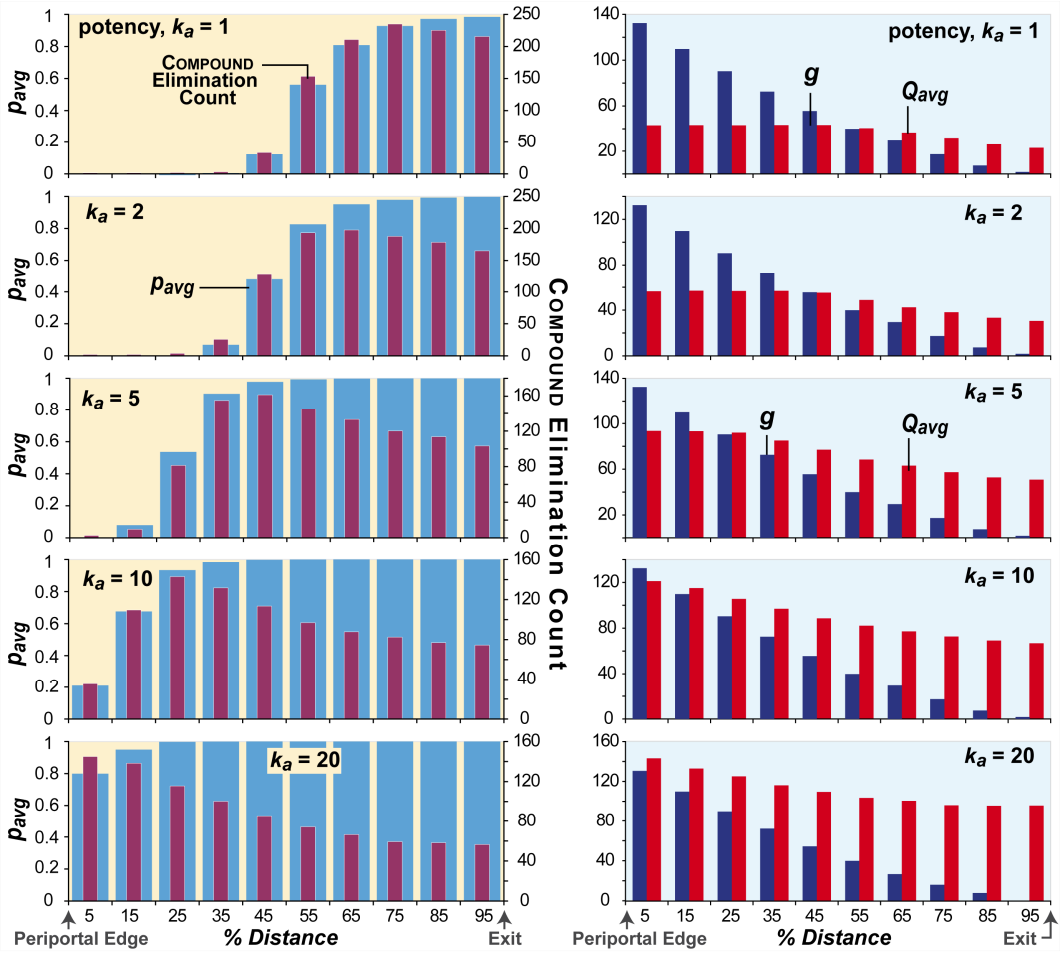

Figure 3.12 Zonation consequences following dosing one ZoRLA with each of five COMPOUNDS having different potencies (k_a). For simplicity, *pBypass* for each SS and each COMPOUND was held constant at 0.8. At the start (t = 0), a random p value was assigned to each SS_i. *A* dose of 50 COMPOUNDS was administered for each of 1,000 simulation cycles. By then, the stable patterns shown had emerged. A simulation cycle continued until all COMPOUNDS exited or were cleared. Values at early t during repeat experiments were not identical, but by t = 1,000 they became identical. X-axis: the distance from PERIPORTAL edge to CENTRAL VEIN exit was subdivided into ten regions. For each k_a experiment, there are four bar graphs. The two on the left provide measures of zonation. p_{avg} is the mean of all p_i's in one of the ten regions; $p_{i,t}$ was updated using h = 5. Using the default value h = 70, we were unable to clearly demonstrate the COMPOUND elimination shift effect. The clearance effort shift effect became more evident for smaller h values. We selected h = 5 to demonstrate that peak COMPOUND elimination count *can* shift from PERIVENOUS to PERIPORTAL as *ka* increases. Clearance effort, p_{avg}, maps to average intrinsic clearance for a lobular tissue sample taken from the same relative location within a lobule. The cumulative total of COMPOUNDS eliminated by each SS_i was recorded. COMPOUND Elimination Count is the mean of those values for each region. On the right g is plotted, the value of the local gradient (which is the same for each of the five k_a experiments), and the mean, regional Q value (Q_{avg}) at t = 1,000.

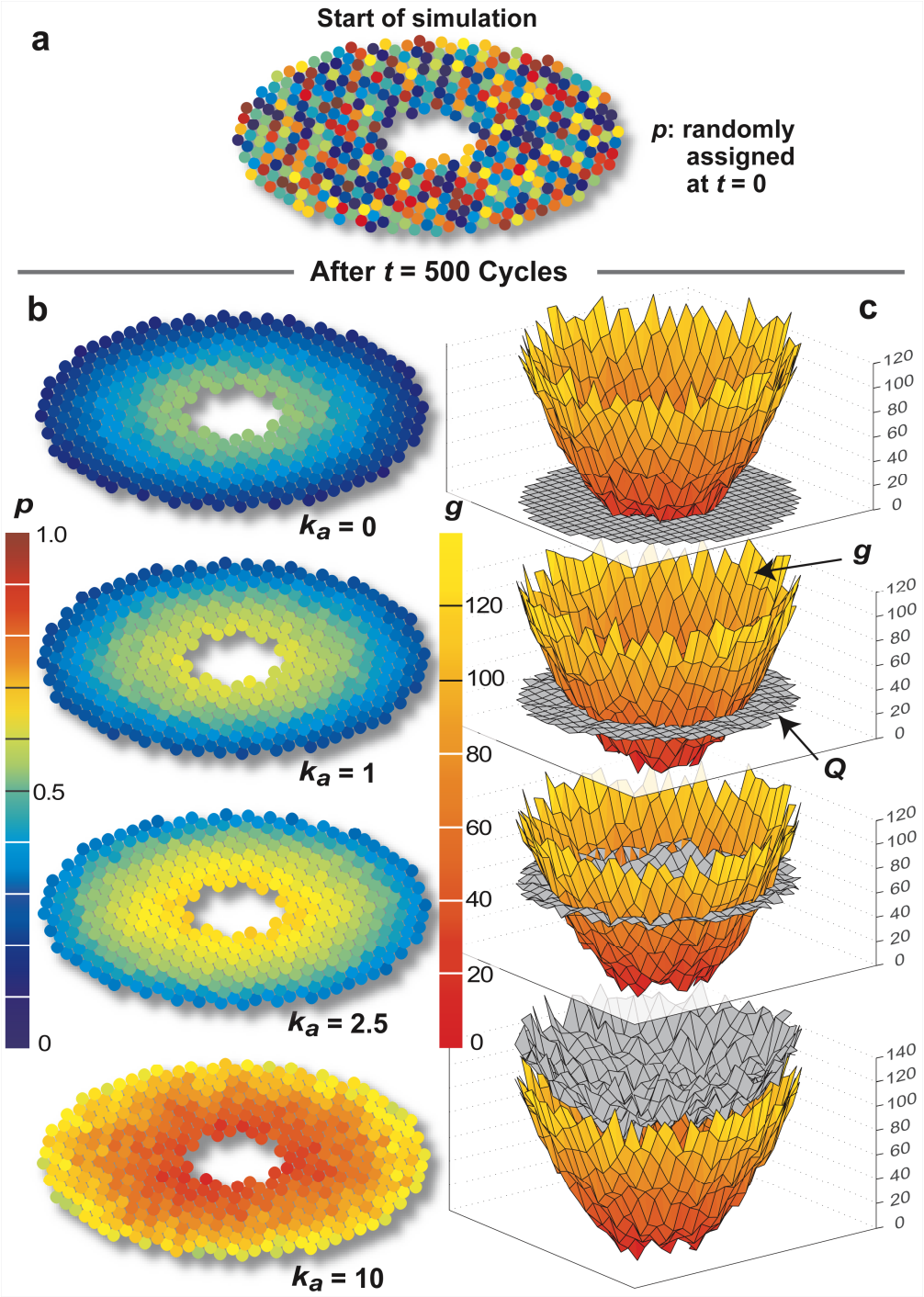

Figure 3.13 ZoRLA zonation patterns. The experiments are the same as for Figure 3.12, except that h = 70. The DOSE each simulation cycle was 50 COMPOUNDS and *pBypass* = 0.8 for all COMPOUNDS. (a) At t = 0, each SS was assigned a random p value (0–1 color scale). (b) The updated value of p_i is shown at the conclusion of the 500th simulation cycle following dosing with one of four COMPOUNDS having different potencies (k_a). (c) Values of g (rust to yellow color scale) and Q_i (gray) are graphed at t = 500.

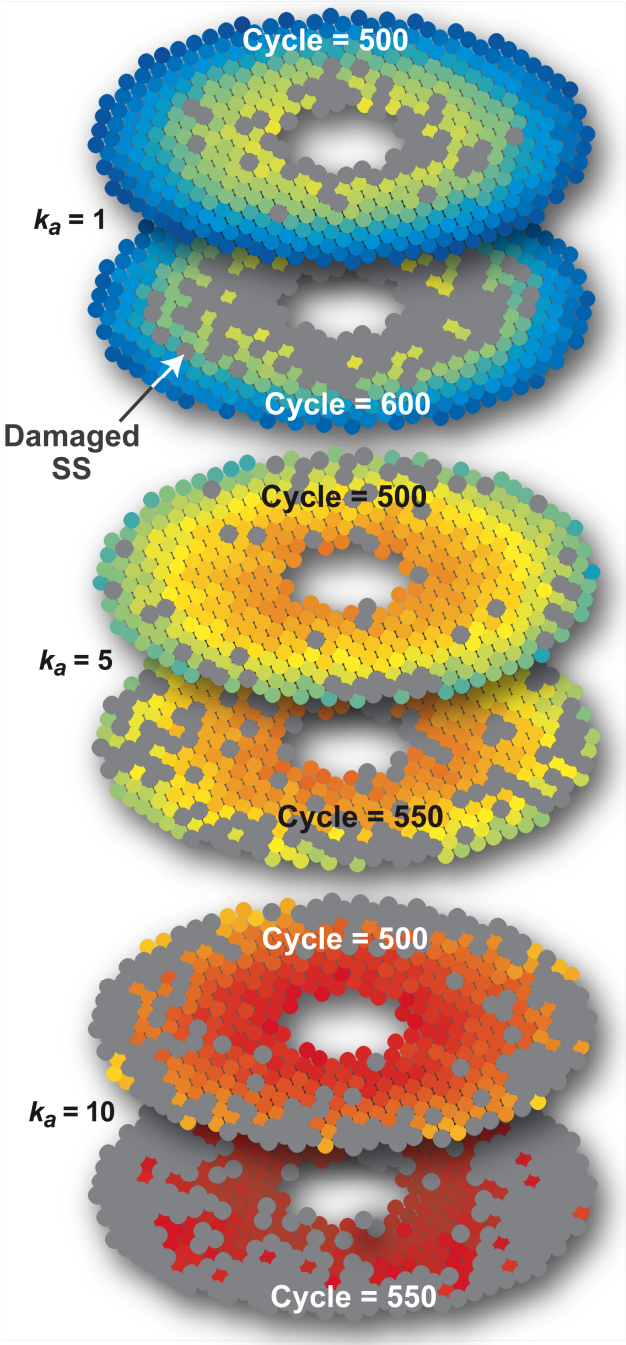

Figure 3.14 Patterns of SS damage. The ZoRLA, experimental protocol, and *p* color scale are the same as in Figure 3.13. We specified that damage (which can map to either severe hepatocyte damage or necrosis) occur after any SS_i CLEARED 50 COMPOUNDS. A damaged SS is gray. Following repeated dosing with one of three COMPOUNDS having potencies of k_a = 1, 5, and 10, the updated value of p_i for an undamaged SS is shown at the conclusion of the 500th and also after the 550th or 600th simulation cycle. The DOSE each simulation cycle was 50 COMPOUNDS.

To explore plausible patterns of HEPATOTOXICITY, we specified that SS damage occur (which can map to either severe hepatocyte damage or necrosis) when any SS in the ZoRLA used for Figure 3.13 clears 50 COMPOUNDS. We conducted experiments for k_a = 0, 1, 2.5, and 10. For k_a = 1, consistent damage is evident after t = 400. It is evident earlier for k_a = 5 and 10. Results are shown in Figure 3.14 at t = 500, t = 550 (k_a = 5 and 10), and t = 600 (k_a = 1). Patterns are similar to the COMPOUND elimination counts in Figure 3.12. The low potency COMPOUND (k_a = 1) caused PERIVENOUS HEPATOTOXICITY, whereas for the higher potency COMPOUNDS (k_a = 5 and 10), there was more PERIPORTAL HEPATOTOXICITY.

Quantitative validation for dose-dependent ZoRLA zonation patterns

The ZoRLA used for Figure 3.13 was reused to explore the consequences of changing the dose per simulation cycle on zonal patterns of clearance effort. The results in Figure 3.15 are for five different repetitive dosing experiments using a medium potency COMPOUND, k_a = 5. The dose per simulation cycle ranged from 10 to 1,000 COMPOUNDS. Shown is the pattern at the conclusion of the 500th cycle. Note that repetitive dosing with 50 COMPOUNDS having k_a = 1 in Figure 3.13 gave a clearance effort zonation pattern quite similar to repetitive dosing with 10 COMPOUNDS having k_a = 5 in Figure 3.15. Note also that repetitive dosing with 50 COMPOUNDS having k_a = 10 in Figure 3.13 gave a clearance effort zonation pattern quite similar to repetitive dosing with 100 COMPOUNDS having k_a = 5 in Figure 3.15.

The data in Figure 3.16a are redrawn from (106). The data show CYP1A2 mRNA expression in periportal and perivenous hepatocytes obtained from rats three days after a single oral dose of 0.01, 0.3 or 10.0 µg [^{3}H]TCDD/kg (2,3,7,8-tetrachlorodibenzo-

p-dioixon); LD_{50} = 0.034 mg/kg, a potent toxin and a potent inducer of P450 isozymes. To what extent can the data in Figure 3.15 be mapped to observations in Figure 3.16a?

We selected a three-step mapping process. First, as done with the wet-lab protocol, we divided the LOBULE into two regions and then averaged the clearance efforts in Figure 3.15 for t = 500. Second, we specified that a simple linear correlation should exist between clearance efforts and log Cyp1A2 mRNA expression levels, and that the levels of induction for the high dose are close to the maximum that can be achieved (they correspond to $p \approx 1.0$). The one that gave the best y-axis match was transformed data = $3.8(p_{avg}) + 4.6$. Those transformed values are graphed in Figure 3.16b. Step three was placement of the transformed ZoRLA data relative to the x-axis (which is [^{3}H] counts—assumed to be TCDD—per viable hepatocyte isolated from periportal and perivenous tissue samples three days after dosing). That mapping was not straightforward. SSs map to more than just hepatocytes. There is not a 1:1 mapping from a COMPOUND to an amount of TCDD. Further, a COMPOUND in low and high dose ZoRLA experiments can (and should) map to different amounts of TCDD. Finally, although we measured average COMPOUND elimination count (CEC_{avg}) for PERIPORTAL and PERIVENOUS SS after 500 simulation cycles, we did not include processes that would map to either 1) elimination and removal of metabolites or 2) redistribution of TCDD and metabolites to tissues. Nevertheless, we sought a single transform that would provide a reasonable quantitative mapping. The one used in Figure 3.16b is transform CEC_{avg} according to the following log-linear equation: $\log(x) = 2 \cdot \log(CEC_{avg}) - 9$.

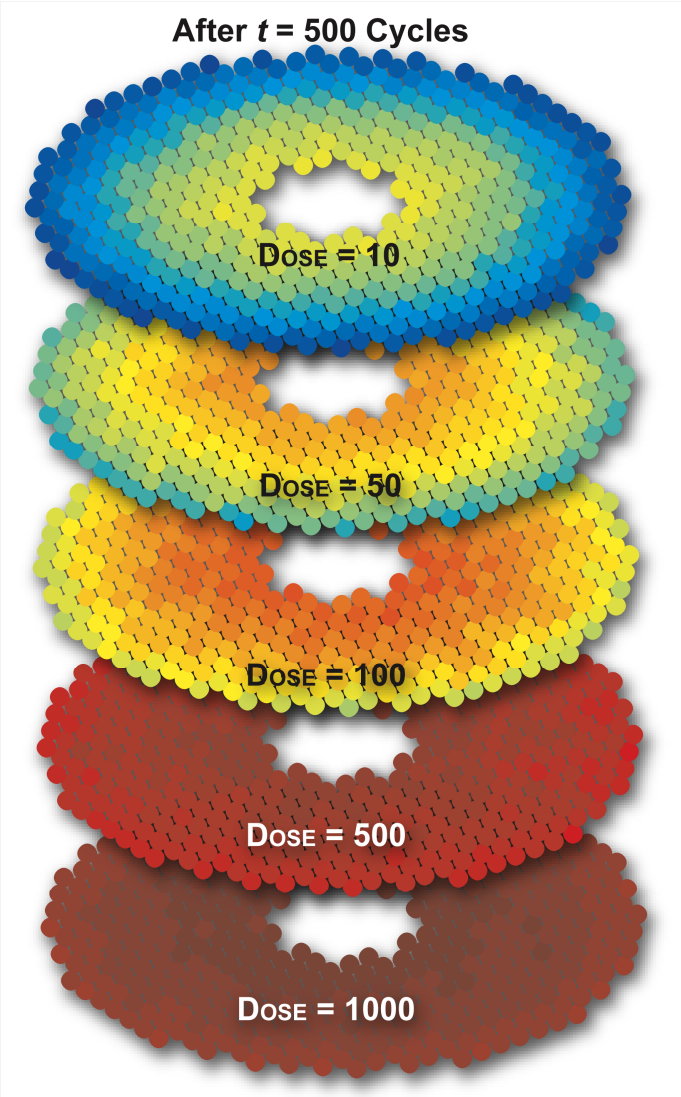

Figure 3.15 Dose-response patterns. The ZoRLA, experimental protocol, and color scale are the same as in Figure 3.13. The damage option is turned off. Five experiments were conducted. Each simulation cycle began with the indicated dose, 10 to 1,000 COMPOUNDS, having $k_a = 5$. Shown is the pattern of p_i values at the conclusion of the 500th simulation cycle. These data are used in Figure 3.16b.

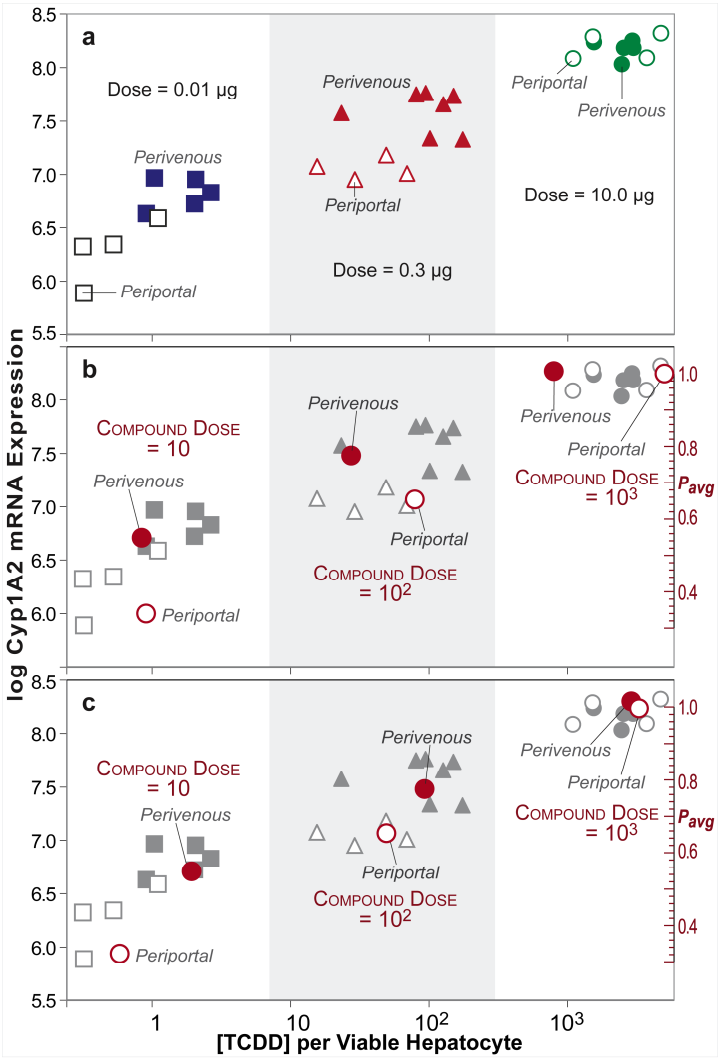

Figure 3.16 Wet-lab and ZoRLA data showing dose-dependent zonation. (a) Dose dependent (x-axis) localization of [^{3}H]TCDD (2,3,7,8-tetrachlorodibenzo-p-dioixon) and CYP1A2 mRNA expression (y-axis) in periportal and perivenous hepatocytes obtained from rats three days after a single oral dose of 0.01, 0.3 or 10.0 µg TCDD/kg; redrawn from (106). X-axis: levels of [^{3}H]TCDD in the two cell subpopulations expressed as attograms [^{3}H]TCDD/viable hepatocyte; y-axis: log Cyp1A2 mRNA expression level. (b) The mean PERIPORTAL and PERIVENOUS clearance effort values (p_{avg}, right axis) from Figure 3.15 are shown mapped to the wet-lab data using the transforms provided in the text. The ZoRLA space was divided into two equal regions (zones) and the mean p_i (p_{avg}: right axis) along with mean COMPOUND elimination count within each region were calculated. The measurements were transformed to map quantitatively to the wet-lab data (both y- and x-axis placement) as described in Results. (c) The experiments that yielded the Figure 3.15 DOSE-response patterns were repeated using *pBypass* = 0.99 rather than 0.8. p_{avg} and mean COMPOUND elimination count were calculated as in (b). The measurements were transformed to map quantitatively to the wet-lab data as described in Results.

Note that in Figure 3.16a, for the two lower doses, there is more TCDD per perivenous than periportal hepatocyte. For ZoRLA experiments, there is little difference in the corresponding low dose transformed CECavg values. For the two larger doses, however, the periportal transformed CECavg values were largest. Changing pBypass can alter that trend. For the experiments that correspond to Figures 3.12–3.15, pBypass was arbitrarily set to 0.8. We increased pBypass to 0.99, which would map to compound that is extensively bound to blood proteins, and repeated the experiments in Figure 3.15. The transformation of pavg to log Cyp1A2 mRNA expression levels (y-axis) was the same as for (b). We again sought a single transform that would provide a reasonable quantitative mapping of CECavg to TCDD per viable hepatocyte. The one used in Figure 3.16c transforms CECavg as follows: $\log(x) = 1.3 \cdot \log(\text{CECavg}) + 0.51$. Note that the similarity between perivenous—periportal differences in transformed CECavg values for each dose and the corresponding wet-lab values has improved.

3.2.4 Discussion

Plausible mappings of SS mechanisms to hepatic counterparts

Braeuning recently presented arguments and supporting evidence for the Ras-Raf-MAPK and Wnt-β-catenin signaling pathways playing roles in both hepatocyte zonation as well as induction of P450 isozymes (97). There are clear consistencies between the ideas presented in that paper and the more abstract preZoRLA1 mechanisms, and they are identified in Appendix C.

Braeuning also presents evidence of an overall deactivating or repressive effect of Ras-Raf-MAPK (mitogen-activated, protein kinase) signaling on expression of CYP

enzymes. Because detection of local gradients (*g*-subsystem) by the CLEARANCE Management Module in Figure 3.10c causes decrement of clearance effort, we conjecture that a mapping exists between that ZoRLA process and processes that include the Ras-Raf-MAPK signaling pathway. In addition, there is evidence for direct transcriptional activation of CYPs by Wnt-β-catenin signaling. The case is made that the β-catenin pathway can be regarded as a decisive factor in the regulation of the perivenous hepatocyte gene expression profile, including basal expression of CYPs and other xenobiotic-metabolizing enzymes. Because detection of response signals by the CLEARANCE Management Module causes increment of clearance effort, we conjecture that a mapping also exists between that process and hepatocyte processes that include the Wnt-β-catenin signaling pathway. With further iterative refinement of the CLEARANCE Management Module, those currently abstract mappings can be made increasingly concrete.

When designing a synthetic analogue, there is a strong inclination to insert counterparts to specific biological features, such as including a component that maps directly to the Wnt-β-catenin signaling pathway, simply because current evidence indicates that the component plays some role in generating the targeted phenomena. Doing so prematurely is a mistake when we are not yet confident of the component's putative role in causing the phenomena. So doing forces establishing groundings to other components at a specific level of granularity that may or may not be warranted for the attributes targeted coupled with the current levels of knowledge, ignorance, and uncertainty. The components in Figure 3.10 are more abstract and coarse-grained than

the Wnt-β-catenin signaling pathway. They were made only as complicated and specific as needed to achieve the patterns and attributes targeted. They can be made more fine-grained and specific iteratively, as the set of attributes targeted expands. Once a degree of validation has been achieved for coarse-grained components, their behaviors during simulation can be used for cross-model validation during refinement to more fine-grained (greater mechanistic detail) counterparts.

The current abstract micromechanisms map to a conflation of all fine-grained processes in lobules that contribute to the simulated events. If we do not have specific evidence on how some known signaling and/or metabolic pathway is contributing, then there is no scientific value in simply implementing some conceptual linkage simply for the sake of including it. So, how can one achieve such linkage?

Consider the following. We have observations from identical experiments on two different livers, one normal (wild type) and another from a mouse that has had one or more components within a signaling &/or metabolic pathway measurably altered, e.g., by genetic deletion. Wet-lab data from the knockout's liver (but not that from the normal liver) falsifies the current micromechanism. That evidence forces us to posit one or more new, more fine-grained micromechanisms that incorporate one or more features containing the knocked-out component. Such an approach was used by Tang and Hunt (122) to falsify a coarse-grained micromechanism and replace it with one more finegrained in which components mapped directly to individual macromolecules.

A related question is, how can one parse specific micromechanisms from the overall, systemic behaviors observed experimentally? An effective strategy is to expand the

variety of systemic attributes targeted. Addition of a new attribute can falsify a micromechanism. To revalidate, it may be necessary to replace a coarse-grained micromechanism with one that is somewhat more finegrained.

Following that, a clearer mapping may exist between the more fine-grained micromechanism and the signaling and/or metabolic pathways of interest. Further iterative refinement can increase the concreteness of the mapping. Such a process was used by Lam and Hunt (98) to move iteratively from coarser to more fine-grained micromechanistic hypotheses.

Inverse maps from phenomena to generators

The exploration of an inverse map from phenomena (patterns of zonation) to generators requires one to hypothesize and then build generators that may cause the phenomena. The question posed is this: given phenomena such as intralobular zonation, what plausible generators might cause their emergence over time? However, a generator–phenomenon map is not one-to-one. Many generator compositions (mechanisms) can produce phenomena that when measured are indistinguishable. Nevertheless, the first step is to find and validate one, which is what we have done here. So doing opens the door to discovering others. When we find a hypothetical generator in the form of an extant biomimetic mechanism, we do not yet have new biological knowledge, but we do have a concrete instance of a strictly defined, plausible and observable mechanism within a system suitable for experimentation, as called for by Christoffels et al. (73), whereas before we only had unchallenged concepts. In the absence of other concrete, competing theories, that system and its mechanism can stand as the current best explanation of the phenomena's cause until falsified by evidence. A

systematic, scientific study of any inverse map (phenomenon-to-generator) can only be done with concretizable hypotheses, either biological or *in silico*. It cannot be done with hypotheses that remain conceptual (2).

As this new M&S method matures, we anticipate that the preferred approach will be to identify several somewhat different yet still plausible generator models and refine them in parallel against an expanding set of targeted attributes. We can expect that modest, selective expansions in the set of targeted attributes will eliminate some generator concepts but allow others to advance following refinement.

Objective achieved

In Methods, we specified that R-SIGNALS are COMPOUND dependent, and listed three mechanistic options: 1) a COMPOUND or its METABOLITE functions as a R-SIGNAL; 2) a COMPOUND or its METABOLITE initiates a response within a SS and that response (or an object generated by that response) functions as a R-SIGNAL that is also detectable subsequently by other SS; and 3) a COMPOUND or its METABOLITE that escapes clearance initiates an EXTRAHEPATIC response, and that response (or an object generated by that response) is an R-SIGNAL that is SS detectable. We implemented the third option, because early exploratory simulations indicated that it would be the simplest to implement. Exploratory simulations (Appendix CFigure C.5 and Figure C.6) showed that zonation does occur when using COMPOUNDS as R-SIGNALS for single-pass dosing experiments. However, further exploration (not shown) demonstrated that the first option could not achieve targeted attribute 7 because the R-SIGNALS generated stay within each SS and so cannot influence other SSs. Option two is more complicated than option three. Given the evidence presented herein, it merits exploration. It seems likely that

there are several mechanisms by which zonation of hepatic cellular properties could occur. The approach used herein can be extended to discover and investigate those options.

Reports on regional hepatic CYP1A1 and CYP1A2 induction describe a "switch-like" hepatocyte response to TCDD: a sharp boundary between areas of induced and noninduced cells (107, 108, 109). Although this phenomena was not among the ten attributes targeted, we discovered that decreasing h to very small values, e.g., 0.01 and smaller, caused ZoRLA to exhibit a clear-cut, switch-like behavior, in which the zonal location of the switch is dose dependent. An example is provided in Figure C.7.

With the 2D ZoRLA design in Figures 3.10 and 3.11, the abstract mechanism in Figure 3.10c, and the results in Figures 3.12 – 3.16, we achieved targeted attributes 1–10, with two exceptions, attributes 8c & d. The effects of hypophysectomy and other treatments on lobular mRNA distributions of various P450 isozymes (attribute 8d) is fascinating, but beyond the scope of this simple ZoRLA. Attribute 8c, cites an example of a xenobiotic inducing zonation patterns that are the reverse of those in Figures 3.13 and 3.15. If g_i in Eq. (3.6) is replaced by $(g_{max} - g_i)$, the patterns in Figures 3.13 and 3.15 are reversed.

Based on the preceding evidence, we suggest that hepatic counterparts to the abstract ZoRLA mechanism, along with the zonation patterns produced, exist in rats when treated with some xenobiotics. The implication of these ZoRLA experiments is that hepatocytes do learn from experience and can cooperate to remove xenobiotics, including some that produce significant response or toxicity.

The results also demonstrate a new scientific method to experimentally explore and

challenge mechanistic hypotheses about the causal nature of toxicologic and pharmacologic phenomena.

Table 3.4 Targeted Attributes and Specifications. Hepatic lobules are the referent. The model system is called a zonally responsive lobular analogue (ZoRLA). **A:** a targeted attribute; **S:** a ZoRLA specification. Attributes 1–10 are targeted for this report.

Attribute Number	**A: a targeted attribute** **S: a ZoRLA specification**
1	**A**: Lobules are comprised of cells, which are typically autonomous functional units. **S**: Each ZoRLA is comprised of autonomous functional units called sinusoidal segments (SS) that map to a portion of a sinusoid containing one or more hepatocytes (Figure 3.10).
2	**A**: Hepatocyte function is location dependent. Cells respond based only on local information. **S**: A SS does the same.
3	**A**: Upon interaction with hepatic cells, blood-borne material (such as O_2), which function as signals, form periportal-to-perivenous (P-to-P) gradients. Those gradients provide information to hepatocytes about their relative location. **S**: A SS should be capable of forming and responding to various trans-lobular signal gradients.
4	**A**: Upon administration via blood, compounds, alone or in combination, having different potencies and different affinities for hepatic functions, percolate through lobules exhibiting P-to-P movement. **S**: Mobile objects carrying identification information map to small fractions of a xenobiotic dose. When a compound generates a response, there is a ZoRLA counterpart to potency. It must be easy to change compound dose, potency, movement, and affinity characteristics.
5	**A**: Hepatocytes clear xenobiotics from blood (metabolism and biliary elimination). A standard measure hepatic removal of a xenobiotic is intrinsic clearance (CLint), the clearance rate absent blood flow. It depends in part on the relative amounts of enzymes and transporters expressed by each hepatocyte. Location-dependent CLint can change following exposure to a xenobiotic. **S**: The ZoRLA counterpart to CLint is clearance effort. Each SS can change its clearance effort upon exposure to compounds.
6	**A**: Hepatocytes often exhibit location dependent expression of enzymes and transporters. **S**: SSs can exhibit location dependent clearance effort.

Attribute Number	A: a targeted attribute S: a ZoRLA specification
7	**A**: Zonation of xenobiotic metabolism occurs at both the cell and lobule level, implicating cell-cell cooperation at some level. It changes adaptively. **S**: Individual SSs can improve efficiency by adapting their clearance effort to changing compound exposures and to actions of other SS.
8	**A**: Xenobiotic dosing, using a variety of experimental conditions, produces different P-to-P induction patterns for different members of the CYP family. Examples (provided for convenience in Appendix C, Figure C.10) include: Induction patterns of CYP1A1, -1A2 and -1B1 by TCDD (106, 110, 111) Induction of CYP2B1/2 by octamethylcyclotetrasiloxane (112) The CYP1A1 mRNA induction pattern caused by β-naphthoflavone is opposite to that caused by 3-methylcholanthrene (113) Effects of hypophysectomy, growth hormone, and tri-iodothyronine on the lobular mRNA distribution of various CYP forms (90) (not yet achieved) Perivenous induction of CYP2A1, -2B1, -2C11, -2E1 and -3A1 by acetone, ethanol or phenobarbital (114) Periportal induction of CYPEtOH2 by ethanol or phenobarbital (114) (not yet achieved) Dose-dependent zonation of enzyme induction: perivenous induction with low-dose, and periportal induction with high-dose phenobarbital treatment (114) **S**: A ZoRLA should be capable of producing patterns of clearance effort that mimic those observed *in vivo* over variety of experimental conditions. A degree of validation is achieved by using a simple transformation to achieve quantitative similarity.
9	**A**: Most histological zonation data is in the form of stained, 2D sections through liver samples that also provide evidence of unique sinusoidal flow paths. **S**: ZoRLAs consists of 2D SS arrangements that enable mobile objects to follow multiple, different P-to-P paths.
10	**A**: Compound dosing can cause dose dependent zonal patterns of cell damage. Examples (provided for convenience in Figure C.10) include: Cell necrosis patterns by anterograde infusion of $CBrCl_3$ (115) Perivenous damage by CCl_4 and bromobenzene (116) Periportal damage by digitonin (116) **S**: A ZoRLA can produce dose dependent damage patterns that mimic those above.
11	**A**: When retrograde infusion is used, different clearance/damage patterns emerge (see (115) for a $CBrCl_3$ example). **S**: ZoRLA produce similar patterns when movement of mobile objects is reversed.

Attribute Number	A: a targeted attribute S: a ZoRLA specification
12	**A**: Histological micrographs show zonation patterns extending over several adjacent lobules. **S**: 2D multi-lobular ZoRLA arrangements that mimic such cross-sections produce similar patterns of clearance effort.
13	**A**: Cell response depends in part on the physicochemical properties of compounds. **S**: A ZoRLA is able to use physicochemical property information of a compound to which a mobile object maps in place of a potency value to produce zonation patterns that cross-validate.
14	**A**: Several liver fractions exhibit zonation concurrently (e.g. protein synthesis, fatty acid degradation, xenobiotic metabolism, etc.). **S**: ZoRLA zonation mechanisms are scalable to produce concurrent zonation of various functions.

4 Summary and Conclusions

Computational approaches for discovering mechanisms that influence treatment responses are evolving rapidly. New software engineering technologies provide researchers with computational means to exploit ideas and evaluate hypotheses more productively. I proposed a novel mechanistic explanation of observed tissue level patterns of liver detoxification and compound-induced toxicity based on simulations using multi-agent models. In the models presented herein, autonomous agents map to collections of liver cells. Individual behaviors at the agent level were specified. Upon parameterization, agent-level actions resulted in a collective behavior similar to that observed for toxin elimination within hepatocyte cultures and liver lobules.

I followed a relatively new scientific modeling approach, called the synthetic method of modeling and simulation (2). Utilizing recent advances in computer technology, synthetic modeling can accelerate biological knowledge discovery by facilitating fast-paced cycles of hypothesis generation, selection, and falsification. Mechanistic hypotheses are created by abductive reasoning (2). Successful mechanistic hypotheses can suggest new experiments to more deeply probe the phenomena of interest (wet-lab

and *in silico*). Scientific modeling and simulation requires use of all three reasoning methods: abduction, induction, and deduction.

In particular, I presented two different models of hepatic xenobiotic elimination[1]. The first, *in silico* hepatocyte (Chapter 2), focuses on cultured primary hepatocytes. The goal for this model is predicting *in vitro* drug clearance. The traditional approach to predicting the *in vitro* and *in vivo* properties of new compounds is to search for patterns within large data sets of measured biological property data and then seek patterns within the set of compound property values of those compounds for which correlations exist. Knowledge about the mechanisms that generated the biological data is only used indirectly. I combined the knowledge of mechanisms and the patterns found in the space of the physicochemical and biological properties. The models and approach are designed to leverage that knowledge by representing and improving our understanding of the generative relationships within the target biological system. The approach is new: more work is needed to uncover and understand limitations and to delineate advantages relative to other methods (optimization, data fitting, etc).

In Chapter 3, I presented mathematical, cost-based, multi-agent simulations of hepatic xenobiotic clearance processes. The models consist of a group of agents that, similar to hepatocytes, cooperate to protect the system of which they are a part (the organism) against toxins. Agents did not have a priori information about either the environment or other agents. Each agent performed *Q*-learning independently. *Q*-learning is a primitive form of learning that minimizes the agent's long-term discounted costs.

[1] All models developed in this work were constructed using MATLAB. The code is available online at biosystems.ucsf.edu.

Costs to an agent were specified to be proportional to the agent's energy expenditure associated with toxin elimination, and the damage caused to the organism due to toxic effects. The collective behavior of those agents was strikingly similar to that of hepatic cells within lobules.

Adaptive cooperation among hepatic cells may have been an important factor in the evolution of within organ, location-dependent strategies used by cells to deal with short- and long-term changes in each cell's environment as well as that of the organism. Simulations of the type used provide new insight into such processes, which in turn improves our ability to anticipate the metabolic and biological fate of compounds of interest. Doing so is expected to expedite drug discovery and development.

5 References

1. Schichl H. Models and the History of Modeling. Kallrath J, editor. Modeling Languages in Mathematical Optimization. Kluwer Academic Publishers. Norwell, MA, USA. 2004:25-36.
2. Hunt CA, Ropella GEP, Lam TN, Tang J, Kim SHJ, Engelberg JA, Sheikh-Bahaei, S. At the Biological Modeling and Simulation Frontier. Pharm Res. 2009; 26(11):2369-400.
3. Zeigler BP, Kim TG, Praehofer H. Theory of Modeling and Simulation. Academic Press, Inc. Orlando, FL, USA. 2000.
4. Ropella GEP, Hunt CA, Nag DA. Using Heuristic Models to Bridge the Gap between Analytic and Experimental Models in Biology. Proceedings of the Agent-Directed Simulation Symposium (San Diego, CA, Apr 2-8). 2005:182-90.
5. Yan L, Hunt CA, Ropella GEP, Roberts MS. In silico Representation of the Liver-Connecting Function to Anatomy, Physiology and Heterogeneous Microenvironments. Conf Proc IEEE Eng Med Biol Soc. 2004; 2:853-6.
6. Hunt CA, Ropella GEP, Roberts MS, Yan L. Biomimetic In Silico Devices. Lect Notes Comput Sc. Springer; 2005;3082:34-42.
7. Gebhardt R. Metabolic Zonation of the Liver: Regulation and Implications for Liver Function. Pharmacol Ther. 1992;53(3):275-354.
8. Roberts MS, Anissimov YG. Modeling of Hepatic Elimination and Organ Distribution Kinetics with the Extended Convection-Dispersion Model. J Pharmacokin. Biopharm. 1999; 27(4):343-82.
9. Yan L, Ropella GEP, Park S, Roberts MS, Hunt CA. Modeling and Simulation of Hepatic Drug Disposition using a Physiologically Based, Multi-Agent In Silico Liver. Pharm Res. 2008;25(5):1023-36.
10. Ropella GEP, Park S, Hunt CA. Evaluating an Hepatic Enzyme Induction Mechanism through Coarse-and Fine-Grained Measurements of an In Silico Liver. Complexity. 2008;14(6):28-34.

11. Sheikh-Bahaei S, Ropella GEP, Hunt CA. In Silico Hepatocyte: Agent-Based Modeling of the Biliary Excretion of Drugs In Vitro. Simulation Series. 2006;38(2):157-63.

12. Sheikh-Bahaei S, Hunt CA. Prediction of In Vitro Hepatic Biliary Excretion using Stochastic Agent-Based Modeling and Fuzzy Clustering. Proceedings of the 38th Winter Simulation Conference. 2006:1617-24.

13. Sheikh-Bahaei S, Ropella GEP, Hunt CA. Hu G, editor. Parameter Estimation Via Analysis of Fuzzy Clusters (PEAF): An Algorithm to Estimate Parameters of Agent-Based Models. Proceedings of the 20th International Conference on Computer Applications in Industry and Engineering; Nov 7-9; San Francisco. Cary, NC: ISCA; 2007:3-7.

14. Abbott R. Emergence Explained: Abstractions: Getting Epiphenomena to Do Real Work. Complexity. 2006;12(1):13-26.

15. Shalizi CR. Methods and Techniques of Complex Systems Science: An Overview. Deisboeck TS and Kresh JY, editors. Complex Systems Science in Biomedicine. Springer US. 2006:33-114.

16. Arias IM, Alter HJ, Cohen DE, Fausto N, Shaftritz DA, Wolkoff AW. The Liver, Biology and Pathobiology. 5th ed. John Wiley & Sons; 2009.

17. Theraulaz G, Bonabeau E. A Brief History of Stigmergy. Artif Life. 1999;5(2):97-116.

18. Takahashi K, Kaizu K, Hu B, Tomita M. A Multi-Algorithm, Multi-Timescale Method for Cell Simulation. Bioinformatics. 2004;20(4):538.

19. Ropella GEP, Hunt CA, Sheikh-Bahaei S. Methodological Considerations of Heuristic Modeling of Biological Systems. Proceedings of the 9th World Multi-Conference on Systemics, Cybernetics and Informatics. 2005:10-3.

20. Liu X, LeCluyse EL, Brouwer KR, Gan LSL, Lemasters JJ, Stieger B, Meier PJ, Brouwer KL. Biliary Excretion in Primary Rat Hepatocytes Cultured in a Collagen-Sandwich Configuration. Am J Physiol- Gastr L. 1999;277(1):12-21.

21. Gallbladder & Bile Duct Cancer: Anatomy and Physiology of the Gallbladder and Bile Ducts [Homepage on the Internet]. The Johns Hopkins University. Available from: http://pathology2.jhu.edu/gbbd/anatphys.cfm.

22. Ropella GEP, Hunt CA. Prerequisites for Effective Experimentation in Computational Biology. Proceedings of the 25th Annual International Conference of the Engineering in Medicine and Biology Society (Cancun, Sept 17-21); 2003:1272-5.

23. Sheikh-Bahaei S, Ropella GEP, Hunt CA. Agent-Based Simulation of In Vitro Hepatic Drug Metabolism: In Silico Hepatic Intrinsic Clearance. Proceedings of the 2005 Agent-Directed Simulation Symposium (San Diego, CA, apr 2-8), Simulation Series. 2005;37(2):171-6.

24. Liu Y, Anthony Hunt C. Studies of Intestinal Drug Transport using an In Silico Epithelio-Mimetic Device. BioSystems. 2005;82(2):154-67.

25. Liu Y, Hunt CA. Mechanistic Study of the Interplay of Intestinal Transport and Metabolism using the Synthetic Modeling Method. Pharm Res. 2006;23(3):493-505.

26. Ropella GEP, Nag DA, Hunt CA. Similarity Measures for Automated Comparison of In Silico and In Vitro Experimental Results. Proceedings of the 25th Annual International Conference of the Engineering in Medicine and Biology Society (Cancun, Sept 17-21). 2003:2933-6.

27. Nelder JA, Mead R. A Simplex Algorithm for Function Minimization. Comput J. 1965;7(4):308-13.

28. Luersen MA, Le Riche R. Globalized Nelder–Mead Method for Engineering Optimization. Comp Struct. 2004;82(23-26):2251-60.

29. Tan SYGL, Van Oortmarssen GJ, Piersma N. Estimating Parameters of a Microsimulation Model for Breast Cancer Screening using the Score Function Method. Ann Oper Res. 2003;119(1):43-61.

30. Chelouah R, Siarry P. Genetic and Nelder–Mead Algorithms Hybridized for a More Accurate Global Optimization of Continuous Multiminima Functions. Eur J Oper Res. 2003;148(2):335-48.

31. Lagarias JC, Reeds JA, Wright MH, Wright PE. Convergence Properties of the Nelder-Mead Simplex Method in Low Dimensions. SIAM J Optimiz. 1999;9(1):112-47.

32. Neddermeijer HG, van Oortmarssen GJ, Piersma N, Dekker R, Habbema JDF. Adaptive Extensions of the Nelder and Mead Simplex Method for Optimization of Stochastic Simulation Models. Econometric Institute Report. 2000;199.

33. Augen J. The Evolving Role of Information Technology in the Drug Discovery Process. Drug Discov Today. 2002;7(5):315-23.

34. Rajasethupathy P, Vayttaden SJ, Bhalla US. Systems Modeling: A Pathway to Drug Discovery. Curr Opin Chem Biol. 2005;9(4):400-6.

35. Johnson PE. Simulation Modeling in Political Science. Am Behav Sci. 1999;42(10):1509.

36. Macy MW, Willer R. FROM FACTORS TO ACTORS: Computational Sociology and Agent-Based Modeling. Annu Rev Sociol. 2002;28(1):143-66.

37. Cederman LE. Computational Models of Social Forms: Advancing Generative Process Theory. Am J Sociol. 2005;110(4):864-93.

38. Pepinsky TB. From Agents to Outcomes: Simulation in International Relations. Eur J Intern Relat. 2005;11(3):367.

39. Tesfatsion L, Judd K. Handbook of Computational Economics. Elsevier; 2006.

40. Reynolds CW. Flocks, Herds and Schools: A Distributed Behavioral Model. Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques; ACM New York, NY, USA; 1987:25-34.

41. Scheutz M, Schermerhorn P. Many is More but Not Too Many: Dimensions of

Cooperation of Agents with and without Predictive Capabilities. Proceedings of IEEE/WIC IAT; 2003.

42. Trianni V, Labella TH, Dorigo M. Evolution of Direct Communication for a Swarm-Bot Performing Hole Avoidance. Lect Notes Comput Sc. 2004;3172:130-41.
43. Schermerhorn P, Scheutz M. Implicit Cooperation in Conflict Resolution for Simple Agents. Agent. 2003.
44. Schermerhorn P, Scheutz M. The Effect of Environmental Structure on the Utility of Communication in Hive-Based Swarms. Proceedings of IEEE Swarm Intelligence Symposium; 2005:440-3.
45. Andrews SS, Bray D. Stochastic Simulation of Chemical Reactions with Spatial Resolution and Single Molecule Detail. Phys Biol. 2004;1(3):137-51.
46. Shimizu TS. The Spatial Organisation of Cell Signalling Pathways-a Computer-Based Study [dissertation]. University of Cambridge; 2002.
47. Grimm V. Ten Years of Individual-Based Modelling in Ecology: What have we Learned and what could we Learn in the Future? Ecol Model. 1999;115(2):129-48.
48. Railsback SF, Harvey BC, Lamberson RH, Lee DE, Claasen NJ, Yoshihara S. Population-Level Analysis and Validation of an Individual-Based Cutthroat Trout Model. Nat Resour Model. 2002;15(1):83-110.
49. Scheutz M, Schermerhorn P. Predicting Population Dynamics and Evolutionary Trajectories Based on Performance Evaluations in Alife Simulations. Proceedings of the 2005 Conference on Genetic and Evolutionary Computation; ACM New York, NY, USA; 2005:35-42.
50. Berry BJL, Kiel LD, Elliott E. Adaptive Agents, Intelligence, and Emergent Human Organization: Capturing Complexity through Agent-Based Modeling. P Natl A Sci. 2002;99(90003):7187-8.
51. Conte R. Agent-Based Modeling for Understanding Social Intelligence. P Natl A Sci. 2002;99(90003):7189-90.
52. Hunt CA, Ropella GEP, Yan L, Hung DY, Roberts MS. Physiologically Based Synthetic Models of Hepatic Disposition. J Pharmacokin Pharmacodyn. 2006;33(6):737-72.
53. Grant MR, Mostov KE, Tlsty TD, Hunt CA. Simulating Properties of In Vitro Epithelial Cell Morphogenesis. PLoS Comput Biol. 2006;2(10):e129.
54. Qutub AA, Hunt CA. Glucose Transport to the Brain: A Systems Model. Brain Res Rev. 2005;49(3):595-617.
55. Calvez B, Hutzler G. Parameter Space Exploration of Agent-Based Models. Lect Notes Comput SC. 2005;3684:633.
56. Bezdek JC, Ehrlich R, Full W. FCM: The Fuzzy c-Means Clustering Algorithm. Comp Geosci. 1984;10(2-3):191-203.
57. Pal NR, Bezdek JC. On Cluster Validity for the Fuzzy c-Means Model. IEEE Trans

Fuzzy Syst. 1995;3(3):370-9.

58. McGinnity DF, Soars MG, Urbanowicz RA, Riley RJ. Evaluation of Fresh and Cryopreserved Hepatocytes as In Vitro Drug Metabolism Tools for the Prediction of Metabolic Clearance. Drug Metab Disposition. 2004;32(11):1247-53.

59. Smith DA, Ackland MJ, Jones BC. Properties of Cytochrome P450 Isoenzymes and their Substrates Part 2: Properties of Cytochrome P450 Substrates. Drug Discov Today. 1997;2(11):479-86.

60. Lewis DFV, Eddershaw PJ, Dickins M, Tarbit MH, Goldfarb PS. Structural Determinants of Cytochrome P450 Substrate Specificity, Binding Affinity and Catalytic Rate. Chem Biol Interact. 1998;115(3):175-99.

61. Lewis DFV, Modi S, Dickins M. Structure-Activity Relationship for Human Cytochrome P450 Substrates and Inhibitors. Drug Metab Rev. 2002;34(1-2):69-82.

62. Didziapetris R, Japertas P, Avdeef A, Petrauskas A. Classification Analysis of P-Glycoprotein Substrate Specificity. J Drug Target. 2003;11(7):391-406.

63. Artursson P, Palm K, Luthman K. Caco-2 Monolayers in Experimental and Theoretical Predictions of Drug Transport. Adv Drug Deliv Rev. 1996;22(1-2):67-84.

64. Lau YY, Sapidou E, Cui X, White RE, Cheng KC. Development of a Novel In Vitro Model to Predict Hepatic Clearance using Fresh, Cryopreserved, and Sandwich-Cultured Hepatocytes. Drug Metab Dispos. 2002;30(12):1446-54.

65. Schneider G, Coassolo P, Lave T. Combining In Vitro and In Vivo Pharmacokinetic Data for Prediction of Hepatic Drug Clearance in Humans by Artificial Neural Networks and Multivariate Statistical Techniques. J Med Chem. 1999;42(25):5072-6.

66. Wu CY, Benet LZ. Predicting Drug Disposition Via Application of BCS: Transport/absorption/elimination Interplay and Development of a Biopharmaceutics Drug Disposition Classification System. Pharm Res. 2005;22(1):11-23.

67. Ulanowicz RE, Hannon BM. Life and the Production of Entropy. Proc R Soc Lond B. 1987;232(1267):181-92.

68. Sheikh-Bahaei S, Kim SHJ, Sheikhbahahaei S, Hunt CA. Understanding the Role of Liver Zonation in Toxin Elimination. Int J Intell Control Syst. 2009;14(1):33-40.

69. Sheikh-Bahaei S, Kim SHJ, Hunt CA. Multi-Agent Based Modeling of Liver Detoxification. Proceedings of the 2009 Spring Simulation Multiconference; Society for Computer Simulation International; 2009:1-5.

70. Jungermann K. Zonation of Metabolism and Gene Expression in Liver. Histochem Cell Biol. 1995;103(2):81-91.

71. Abu-Zahra TN, Pang KS. Effect of Zonal Transport and Metabolism on Hepatic Removal: Enalapril Hydrolysis in Zonal, Isolated Rat Hepatocytes In Vitro and Correlation with Perfusion Data. Drug Metab Dispos. 2000;28(7):807-13.

72. Gray MR, Tam YK. The Series-Compartment Model for Hepatic Elimination. Drug Metab Dispos. 1987;15(1):27-31.

73. Christoffels VM, Sassi H, Ruijter JM, Moorman AF, Grange TL,W. H. A Mechanistic Model for the Development and Maintenance of Portocentral Gradients in Gene Expression in the Liver. Hepatology. 1999;29(4):1180-92.

74. Daskalakis C, Goldberg PW, Papadimitriou CH. The Complexity of Computing a Nash Equilibrium. New York, N.Y.: Association for Computing Machinery; 2006:71-8.

75. Daskalakis C, Papadimitriou CH. Three-Player Games are Hard. Electron Colloq Comput Complexity. 2005;139. Available from: http://eccc.hpi-web.de/eccc/.

76. Watkins, C. J. C. H., Dayan P. Technical Note: Q-Learning. Mach Learn. 1992;8(3):279-92.

77. Hu J, Wellman MP. Nash q-Learning for General-Sum Stochastic Games. J Mach Learn Res. 2003;4:1039-69.

78. Yang E, Gu D. Multiagent Reinforcement Learning for Multi-Robot Systems: A Survey. Technical Report CSM-404, Dept. of Computer Science, University of Essex, 2004.

79. Shoham Y, Powers R, Grenager T. If Multi-Agent Learning is the Answer, What is the Question? Artif Int. 2007;171(7):365-77.

80. Liu S, Salyapongse AN, Geller DA, Vodovotz Y, Billiar TR. Hepatocyte Toll-Like Receptor 2 Expression In Vivo and In Vitro: Role of Cytokines in Induction of Rat TLR2 Gene Expression by Lipopolysaccharide. Shock. 2000;14(3):361-5.

81. Hosseini-Yeganeh M, McLachlan AJ. Physiologically Based Pharmacokinetic Model for Terbinafine in Rats and Humans. Antimicrob Agents Ch. 2002; 46(7): 2219.

82. Kandel ER, Schwartz JH. Molecular Biology of Learning: Modulation of Transmitter Release. Science. 1982;218(4571):433-43.

83. Kandel ER, Schwartz JH, Jessell TM. Essentials of Neural Science and Behavior. Appleton & Lange; 1995.

84. Jones BE, Czaja MJ. Intracellular Signaling in Response to Toxic Liver Injury. Am J Physiol Gastrointest Liver Physiol. 1998;275(5):874-8.

85. Xu Y, Jones BE, Neufeld DS, Czaja MJ. Glutathione Modulates Rat and Mouse Hepatocyte Sensitivity to Tumor Necrosis Factor α Toxicity. Gastroenterology. 1998;115(5):1229-37.

86. Herrmann S, Seidelin M, Bisgaard HC, Vang O. Indolo [3, 2-b] Carbazole Inhibits Gap Junctional Intercellular Communication in Rat Primary Hepatocytes and Acts as a Potential Tumor Promoter. Carcinogenesis. 2002;23(11):1861.

87. Dupont G, Tordjmann T, Clair C, Swillens S, Claret M, Combettes L. Mechanism of Receptor-Oriented Intercellular Calcium Wave Propagation in Hepatocytes. FASEB J. 2000;14(2):279-89.

88. Serrière V, Berthon B, Boucherie S, Jacquemin E, Guillon G, Claret M, Tordjmann T. Vasopressin Receptor Distribution in the Liver Controls Calcium Wave Propagation

and Bile Flow. FASEB J. 2001;15(8):1484-6.

89. Gebhardt R, Gaunitz F. Cell-Cell Interactions in the Regulation of the Expression of Hepatic Enzymes. Cell Biol Toxicol. 1997;13(4):263-73.

90. Oinonen T, Lindros KO. Zonation of Hepatic Cytochrome P-450 Expression and Regulation. Biochem J. 1998;329(Pt 1):17-35.

91. Camp JP, Capitano AT. Induction of Zone-Like Liver Function Gradients in HepG2 Cells by Varying Culture Medium Height. Biotechnol Prog. 2007;23(6):1485-91.

92. Jungermann K, Kietzmann T. Oxygen: Modulator of Metabolic Zonation and Disease of the Liver. Hepatology. 2000;31(2):255-60.

93. Lindros K0. Zonation of Cytochrome P450 Expression, Drug Metabolism and Toxicity in Liver. Gen Pharmacol. 1997;28(2):191-6.

94. Benhamouche S, Decaens T, Perret C, Colnot S. Wnt/β-Catenin Pathway and Liver Metabolic Zonation: A New Player for an Old Concept. Med Sci. 2006;22(11):904-6.

95. Sekine S, Lan BYA, Bedolli M, Feng S, Hebrok M. Liver-Specific Loss of b-Catenin Blocks Glutamine Synthesis Pathway Activity and Cytochrome p450 Expression in Mice. Hepatology. 2006;43(4):817-25.

96. Burke ZD, Reed KR, Phesse TJ, Sansom OJ, Clarke AR, Tosh D. Liver Zonation Occurs through a β-Catenin–Dependent, c-Myc–Independent Mechanism. Gastroenterology. 2009;136(7):2316-24.

97. Braeuning A. Regulation of Cytochrome P450 Expression by Ras-and-Catenin-Dependent Signaling. Curr Drug Metab. 2009;10(2):138-58.

98. Lam TN, Hunt CA. Discovering Plausible Mechanistic Details of Hepatic Drug Interactions. Drug Metab Disposition. 2009;37(1):237-46.

99. Darden L. Strategies for Discovering Mechanisms: Schema Instantiation, Modular Subassembly, forward/backward Chaining. Phil Sci. 2002;69(S3):354-65.

100. An G, Mi Q, Dutta-Moscato J, Vodovotz Y. Agent-Based Models in Translational Systems Biology. Wiley Interdiscip Rev Syst Biol Med. 2009;1(2):159-71.

101. Fisher J, Henzinger TA. Executable Cell Biology. Nat Biotechnol. 2007;25(11):1239-50.

102. Grimm V, Revilla E, Berger U, Jeltsch F, Mooij WM, Railsback SF, Thulke HH, Weiner J, Wiegand T, DeAngelis DL. Pattern-Oriented Modeling of Agent-Based Complex Systems: Lessons from Ecology. Science. 2005;310(5750):987-91.

103. Yan L, Sheihk-Bahaei S, Park S, Ropella GEP, Hunt CA. Predictions of Hepatic Disposition Properties using a Mechanistically Realistic, Physiologically Based Model. Drug Metab Dispos. 2008;36(4):759-68.

104. Park S, Ropella GEP, Kim SHJ, Roberts MS, Hunt CA. Computational Strategies Unravel and Trace how Liver Disease Changes Hepatic Drug Disposition. J Pharmacol Exp Ther. 2009;328(1):294-305.

105. Matzinger P. The Danger Model: A Renewed Sense of Self. Science. 2002;296(5566):301-5.

106. Santostefano MJ, Richardson VM, Walker NJ, Blanton J, Lindros KO, Lucier GW, Alcasey SK, Birnbaum LS. Dose-Dependent Localization of TCDD in Isolated Centrilobular and Periportal Hepatocytes. Toxicol Sci. 1999;52(1):9-19.

107. Andersen ME, Birnbaum LS, Barton HA, Eklund CR. Regional Hepatic CYP1A1 and CYP1A2 Induction with 2, 3, 7, 8-Tetrachlorodibenzo-p-Dioxin Evaluated with a Multicompartment Geometric Model of Hepatic Zonation. Toxicol Appl Pharmacol. 1997;144(1):145-55.

108. Andersen ME, Conolly RB. Mechanistic Modeling of Rodent Liver Tumor Promotion at Low Levels of Exposure: An Example Related to Dose-Response Relationships for 2, 3, 7, 8-Tetrachlorodibenzo-p-Dioxin. Hum Exp Toxicol. 1998;17(12):683-90.

109. Broccardo CJ, Billings RE, Chubb LS, Andersen ME, Hanneman WH. Single Cell Analysis of Switch-Like Induction of CYP1A1 in Liver Cell Lines. Toxicol Sci. 2004;78(2):287-94.

110. Tritscher AM, Goldstein JA, Portier CJ, McCoy Z, Clark GC, Lucier GW. Dose-Response Relationships for Chronic Exposure to 2, 3, 7, 8-Tetrachlorodibenzo-p-Dioxin in a Rat Tumor Promotion Model: Quantification and Immunolocalization of CYP1A1 and CYP1A2 in the Liver. Cancer Res. 1992;52(12):3436-42.

111. Walker NJ, Crofts FG, Li Y, Lax SF, Hayes CL, Strickland PT, Lucier GW, Sutter TR. Induction and Localization of Cytochrome P450 1B1 (CYP1B1) Protein in the Livers of TCDD-Treated Rats: Detection using Polyclonal Antibodies Raised to Histidine-Tagged Fusion Proteins Produced and Purified from Bacteria. Carcinogenesis. 1998;19(3):395-402.

112. Sarangapani R, Teeguarden J, Plotzke KP, McKim Jr JM, Andersen ME. Dose-Response Modeling of Cytochrome P450 Induction in Rats by Octamethylcyclotetrasiloxane. Toxicol Sci. 2002;67(2):159-72.

113. Oinonen T, Saarikoski S, Husgafvel-Pursiainen K, Hirvonen A, Lindros KO. Pretranslational Induction of Cytochrome P 4501 A Enzymes by β-Naphthoflavone and 3-Methylcholanthrene Occurs in Different Liver Zones. Biochem Pharmacol. 1994;48(12):2189-97.

114. Buhler R, Lindros KO, Nordling A, Johansson I, Ingelman-Sundberg M. Zonation of Cytochrome P 450 Isozyme Expression and Induction in Rat Liver. Eur J Biochem. 1992;204(1):407-12.

115. Masuda Y, Yamamori Y. Histological Evidence for Dissociation of Lipid Peroxidation and Cell Necrosis in Bromotrichloromethane Hepatotoxicity in the Perfused Rat Liver. Jpn J Pharmacol. 1991;56(2):143-50.

116. Pink DBS. Hepatic Zonation of Δ^1-Pyrroline-5-Carboxylate Metabolism [dissertation]. St. John's Newfoundland Canada: Memorial University of

Newfoundland; 2001.

117. Derrig RA, Ostaszewski KM. Fuzzy Techniques of Pattern Recognition in Risk and Claim Classification. J Risk Insur. 1995:447-82.

118. Wong CC, Chen CC, Yeh SL. K-Means-Based Fuzzy Classifier Design. Fuzzy systems, 2000. FUZZ IEEE 2000. the Ninth IEEE International Conference on Fuzzy Systems; San Antonio, TX; USA; May 2000:48-52.

119. Dearden R, Friedman N, Russell S. Bayesian Q-Learning. Proceedings of the National Conference on Artificial Intelligence; John Wiley & Sons Ltd; 1998:761-8.

120. Ohno H, Naito Y, Nakajima H, Tomita M. Construction of a Biological Tissue Model Based on a Single-Cell Model: a Computer Simulation of Metabolic Heterogeneity in the Liver Lobule. Artif Life. 2008; 14:3-28.

121. Ierapetritou MG., Georgopoulos PG, Roth CM, Androulakis LP. Tissue-Level Modeling of Xenobiotic Metabolism in Liver: an Emerging Tool for Enabling Clinical Translational Research. Clin. Translat. Sci. 2009; 2:228-237.

122. Tang J, Hunt CA. Identifying the Rules of Engagement Enabling Leukocyte Rolling, Activation, and Adhesion. PLoS Comput. Biol. 2010; 6, e1000681.

123. Hailfinger S, Jaworski M, Braeuning A, Buchmann A, Schwarz M. Zonal Gene Expression in Murine Liver: Lessons from Tumors. Hepatology. 2006; 43:407-14.

Appendix A. Nelder and Mead Simplex Method

The Nelder and Mead algorithm, introduced in (27), is a widely used optimization algorithm. Despite its age it is still the method of choice for many practitioners in the fields of statistics, engineering and the physical and medical sciences because it is straightforward to code and easy to use. Particularly, it has been used widely by researchers for simulation optimization 28, 29, 30, 31). It belongs to a class of methods which do not require derivatives and which are often claimed to be robust for problems with discontinuities or where function values are noisy. This property makes it a good candidate for optimizing the stochastic *in silico* simulations.

There are several different versions and extensions of this optimization algorithm. We used the one described in (32) with minor modifications to optimize the parameter values.

Figure A.1 shows the simplex algorithm employed in this work. $\theta_{best}, \theta_{worst}$ and $\theta_{next\text{-}worst}$ are the best (i.e. has the highest similarity score), worst and next worst vertex of the simplex. There are four basic operations used in this algorithm: reflection, contraction, expansion and shrink (described below). The general heuristic in this search method is to move away from the worst point toward the best.

For optimization of an *n*-dimensional stochastic objective function, the simplex algorithm uses a simplex with *n+1* vertices, and evaluates the objective function in every vertex. Based solely on the ranks of the observed function values in the vertices of the simplex, different steps can be taken, such as reflection, expansion, contracting vertices

or shrinking the simplex, in order to find better vertices.

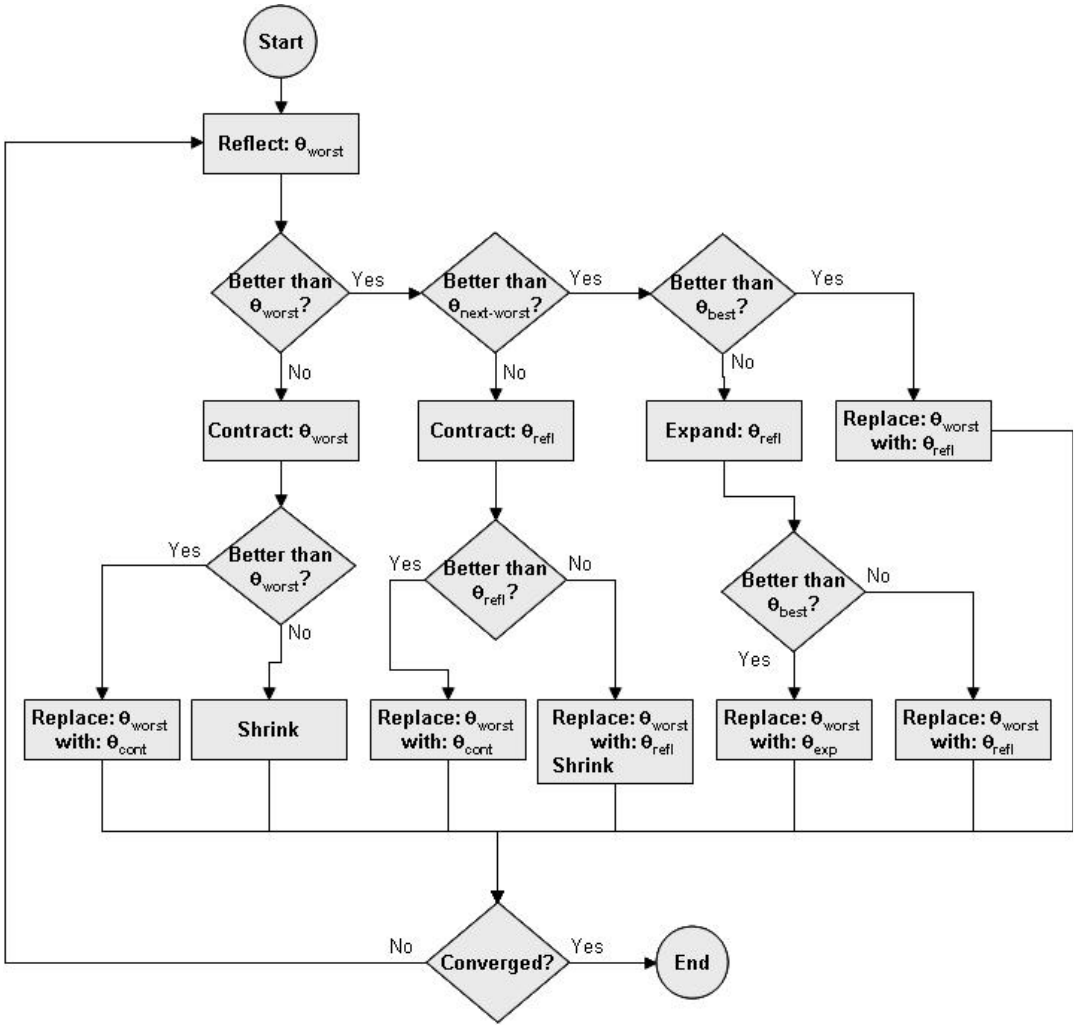

Figure A.1. Nelder and Mead simplex optimization algorithm

In the case of model optimization, the objective function could be a measure of goodness of the model, which in our case is determined by the similarity score assigned to each vertex.

The operations used in the Nelder and Mead Simplex method are:

Reflection: The operation *reflect:* θ *through:* θ_{cent} is defined as:

$$\theta_{refl} = (1+\alpha)\theta_{cent} - \alpha\theta \qquad \alpha > 0$$

where θ_{cent} is the centroid of all simplex vertices except θ; the real number α is called the reflection ratio (Figure A.1a).

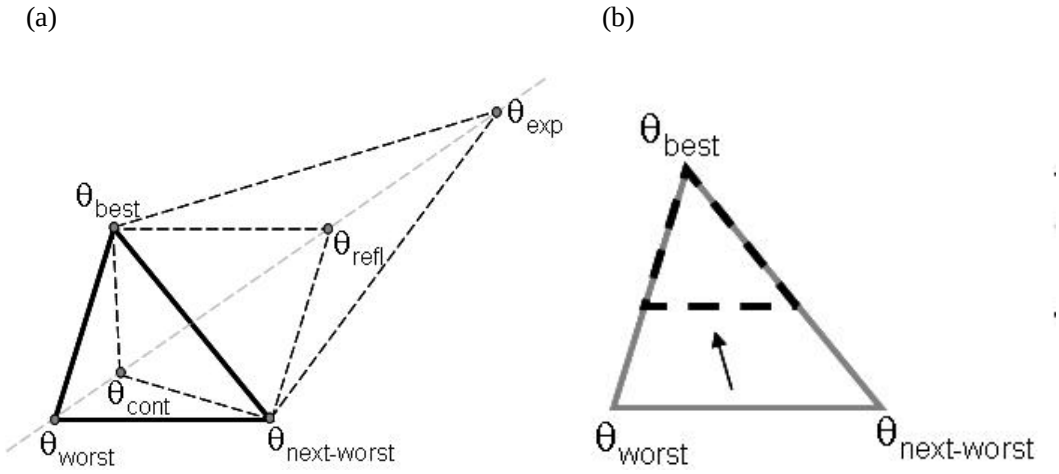

Figure A.2 The simplex four basic operations:
(a) reflect θ_{refl}, contract θ_{cont}, expand θ_{exp} (b) Shrink

Contraction: The operation *contract:* θ *though:* θ_{cent} is defined as:

$$\theta_{cont} = \beta\theta + (1-\beta)\theta_{cent} \qquad 0 < \beta < 1$$

where θ_{cent} is the centroid of all simplex vertices except θ; the real number β is called the contraction ratio (Figure A.1a).

Expansion: The operation *expand:* θ *though:* θ_{cent} is defined as:

$$\theta_{cont} = \gamma\theta + (1-\gamma)\theta_{cent} \qquad \gamma > 1$$

where θ_{cent} is the centroid of all simplex vertices except θ; the real number γ is called the expansion ratio (Figure A.1a).

Shrink: The operation *shrink toward:* θ_{best} is defined as replacing vertex θ_i with

$$\delta\theta_i + (1-\delta)\theta_{best},\ i = 1,...,n+1,\ \theta_i \neq \theta_{best},\ 0 < \delta < 1$$

where δ is a real number between 0 and 1, called the shrink ratio (Figure A.1b).

The Initial Simplex

The simplex method does not start with just one single point but with *n+1* points, defining the initial simplex. One way to form an initial simplex is that if we think of one of these *n+1* points as our starting point θ_0, the other *n* points can be chosen as follows:

$$\theta_{\mathbf{i}} = \theta_{\mathbf{0}} + d\theta_i \times \vec{\mathbf{e}}_{\mathbf{i}}$$

where $\mathbf{e_i}$'s are *n* unit vectors and $d\theta_i$ is our guess of the i^{th} parameter length scale.

Other methods can be used to for an initial simplex as well, for example random selection of the *n+1* points.

As a result the first *n+1* experiments are performed to form the initial simplex and the optimization process begins thereafter.

Appendix B. Fuzzy c-Means Algorithm

Since Fuzzy Set Theory (FST) began in the 60's it has been developed as an alternative to probability theory in modeling uncertainty. Pattern recognition, or search for structure in data, provided the early motivation for developing FST because of the fundamental involvement of human perception (117).

A Fuzzy classifier provides a measure of the degree to which a pattern fits within a class. There are several techniques for Fuzzy pattern recognition. A Fuzzy classifier

based on a genetic algorithm requires a long training time and that time can increase dramatically when the training data has a high dimension. A Fuzzy classifier having an ellipsoidal region has had good performance in many classification problems, but it needs a vast amount of time to calculate the covariance matrix of the ellipsoidal clusters. K-Means-Based Fuzzy classifier (118) uses the K-means algorithm to partition the training data for each class into several clusters, and then some Fuzzy rules are used to construct a Fuzzy classifier. In this work, we use a Fuzzy pattern recognition technique introduced by Bezdek and described in (56): FCM iterative algorithm.

FCM clustering involves minimizing an objective function or error criterion selected from a family of objective function clustering algorithms. A common goal of these algorithms is to find an "optimal" partitioning of feature space given a collection of data samples. The algorithms that, in addition to minimizing an error function, estimate the prototypes of resulting classes within a partition, are often referred to as C-Means clustering algorithms, where the integer *c* stands for the number of classes. If the classes, for which the prototypes are estimated, are allowed to be fuzzy, the FCM clustering algorithm may be used. The FCM algorithm minimizes the least-squares function that is given by a generalized within-groups sum of square errors:

$$J_m = \sum_{k=1}^{n} \sum_{i=1}^{c} \mu_{ik}^m d_{ik}^2$$

where there are *n* observations (in our case the physicochemical properties of the compounds of interest), c classes, μ_{ik} is the membership of x_k (here *x* is the vector of physicochemical properties) in class *i*, *m* (the "Fuzzy exponent") is a parameter used to

control the fuzziness of the class allocation, and

$$d_{ik}^2 = (x_k - v_i)^T G(x_k - v_i)$$

is the distance measure, or inner product norm, between x_k and the mean of class i, denoted v_i, induced by the positive definite weight matrix G, and v_i is determined by

$$v_i = \frac{\sum_{k=1}^{n} \mu_{ik}^m x_k}{\sum_{k=1}^{n} \mu_{ik}^m} .$$

The output of FCM, U, is a real c-by-n matrix, containing the values of the membership functions of the Fuzzy clusters. This matrix satisfies the following two conditions. The first condition is that each feature vector x_k has its total membership value of one divided among all clusters, while the second one states that the sum of membership degrees of feature vectors in a given cluster does not exceed the total number of feature vectors.

Appendix C. Supplement for Section 3.2

C.1 The Linear preZoRLA System

The linear 1D preZoRLA illustrated in Figure C.1C is composed of *n* spatially fixed SS embedded within an extrahepatic space. For the preZoRLA1 and preZoRLA2, *n* = 20. The system maps abstractly to a portion of a hepatic lobule as illustrated in Figure C.1A. SS mechanisms are described below. A SS has no knowledge of any other SS. There are three types of mobile objects: B-SIGNALS, XENOBIOTICS, and R-SIGNALS. A source container (not shown) for each is located just prior to SS_1. The special case

where XENOBIOTICS also function as the R-SIGNALS is allowed. Total SIGNALS are specified by *aDose* and *bDose*. A-SIGNAL dosing can be delayed using the *doseDelay* parameter.

Figure C.1D provides an abstract view of the mechanisms within each SS agent. In a rat lobule, a compound entering a sinusoid segment can exit into the next segment without encountering hepatocytes. The same is true for a SS: a XENOBIOTIC that enters a SS can exit without being detected by (seen by) its CLEARANCE mechanism. The large circle in each SS in Figure C.1D contains that CLEARANCE mechanism. The probability of bypassing a SS is specified by the parameter *pBypass*. A XENOBIOTIC will be detected with a probability of 1 – *pBypass*. Once detected, an event occurs. The XENOBIOTIC either is or is not CLEARED. The primary requirement of the abstract SS CLEARANCE mechanism is that it be consistent with known hepatocyte details. It maps to either detoxification by metabolism (in that case the metabolite is ignored) or elimination into bile. INTRINSIC CLEARANCE of SS_i is the probability p_i that a detected XENOBIOTIC will be CLEARED, and that value is under SS_i control. Because each SS is quasi-autonomous, each can act independently. Increasing p_i can map to induction of metabolizing enzymes and/or induction of uptake or efflux (to bile) transporters.

Simple SS Mechanisms to Create Zonation

Typically, there is no zonation at the start of a simulation. To create measurable, CLEARANCE-related zonation, something within each SS needed to map to metabolic enzymes. To meet that need in preZoRLA1, we specified the PROTEIN objects illustrated in Figure C.2A. PROTEINS can map to any measurable gene product including

xenobiotic metabolizing enzymes. The number (*N*) or use of PROTEINS during a simulation established if zonation had or had not occurred. Each SS also needs a gradient of detectable mobile objects to provide relative location information; SSs working independently should be able to create whatever mobile object gradients are needed. The SS mechanism in Figure C.2 uses two subsystems to create (or activate) and remove (or deactivate) PROTEINS.

SIGNALS, when present in its source container, were metered into a preZoRLA1 at SS_1 at a constant rate, typically one of each type each simulation cycle. Both object types could pass through (bypass) a SS without being detected; the bypass probabilities were *pbBypass* and *paBypass*, for B- and R-SIGNALS, respectively. SIGNAL detection, as illustrated in Figure C.2A, triggered two events in sequence and the SIGNAL was then either removed (with probabilities *pbRemove* and *paRemove*) or released. In the latter case it entered the next SS.

In Figure C.2A, two subsystems regulate the PROTEIN levels. It is too early to speculate on the subcellular networks to which they map. However, we identify candidates in the Introduction. The current subsystems are inductive models serving as placeholders for future sets of concrete interacting components having behaviors that will cross-validate with the current subsystems. A detected B-SIGNAL causes an event in the *b*-subsystem: variable *b* is increased by a constant value k_b; b_t is the average value of variable *b* over N_{avg} simulation cycles ending with the last. For the results presented in Figure C.5, $N_{avg} = 100$. Subscript *t* indexes the value for the current simulation cycle. Subscript *t*+1 indexes the value for the next simulation cycle. The *b* event is followed immediately by another event in the *m*-subsystem: m_t is an integer, proportional to b_t; c_1

is the adjustable proportionality constant. The second event results in removal (or deactivation) of m_t PROTEIN objects.

A detected R-SIGNAL causes an event in the *a*-subsystem: variable *a* is increased by a constant value k_a; a_t is the average value over 100 simulation cycles ending with the last. The a_t event is followed immediately by another event in the *s*-subsystem: s_t is linearly related to a_t and N_t; c_2 and c_3 are tunable constants, which adjust the influence of a_{AVG} and N_{avg} on *s*. Larger c_2 values increase PROTEIN production; c_3 modulates negative feedback on PROTEIN production by already existing PROTEINS, dampening the impact of a change in R-SIGNALS on PROTEIN levels. Independent of all preceding events, a_t and b_t are reduced by one at the end of each simulation cycle. That reduction can map to normal turnover and/or the involvement of those pathways in other cell processes.

Figure C.2B illustrates the mechanism of XENOBIOTIC removal by a SS. A XENOBIOTIC that enters a SS can pass through (bypass) undetected with a probability *pBypass*. The probability of CLEARANCE following detection is p_t. If the XENOBIOTIC is not CLEARED, it will enter the next SS at the start of the next simulation cycle. If CLEARED, it is removed from the simulation. We specified that p_t be a function of N_t (and that it be positively correlated with N_t). When the *N* PROTEINS are responsible for XENOBIOTIC CLEARANCE, $p_t = 1 - (1 - pRemove)^{N_t}$, and *pRemove* is the probability of a XENOBIOTIC being removed by one PROTEIN.

Results of preZoRLA1 operation

Operation of the Figure C.2 mechanism created different patterns of functional zonation across 20 SS connected in series as illustrated in Figure C.1C. One measure of

zonation is the value of *N*, the number of PROTEINS, which can map to the level of expression of a xenobiotic metabolizing enzyme. We explored large regions of preZoRLA1 parameter space and the patterns produced. Six examples are shown in Figure C.5. The first three (Figure C.5A–F) are biomimetic in that the increasing PERIPORTAL-to-PERIVENOUS level of PROTEIN is similar to zonation patterns frequently reported for xenobiotic metabolizing enzymes. Note that b_{avg} values decline approximately linearly from SS_1 to SS_{20}. Detection and removal of B-SIGNALS leads to the equilibrium values of b_{avg}; detection and removal of R-SIGNALS leads to the equilibrium values of a_{avg}. By changing values of c_1–c_3, *paRemove*, and *pbRemove* the mechanism creates a wide variety of PROTEIN level zonation patterns.

Exploring the behavior space of preZoRLA1

PreZoRLA1 is able to exhibit different zonal patterns of PROTEIN expression. The patterns could be categorized into three types: 1) PERIVENOUS expression (Figure C.5A–F); 2) PERIPORTAL expression (Figure C.5G–J); and 3) MID-ZONE expression (Figure C.5 K–L).

To obtain the first behavior, we specified *paRemove* < *pbRemove*. Doing so caused a steeper gradient of B-SIGNAL than that of R-SIGNAL. Consequently, it created a larger PROTEIN expression rate in the PERIVENOUS than that in PERIPORTAL region (Figure C.5B, D, F). The extent of expression can be controlled by parameter c_2. Figure C.5A,C,E shows the results for small, medium and large values of c_2, respectively.

The second behavior was achieved by specifying *paRemove* > *pbRemove*. In this case the R-SIGNAL gradient was steeper than that of B-SIGNAL, and as a result, the rate of

PROTEIN expression was larger PERIPORTAL (Figure C.5H, J). The extent of expression can be controlled by parameter c_2. Figure C.5G, I shows the results for c_2 = 10 and c_2 = 20, respectively.

The third behavior is not easily attained. One example is shown in Figure C.2K. In this example *paRemove* < *pbRemove*; additionally, *pRemove* was greater than zero but very small. Doing so made R-SIGNAL > B-SIGNAL in the MID-ZONAL region; but R-SIGNAL < B-SIGNAL in the other two regions. Unlike the above two situations, a_{AVG} and b_{AVG} values were unstable and kept increasing as long as the simulation was running (Figure C.5L).

The above instability is abiotic. Switching from the somewhat more complicated *Q* learning mechanisms in Figure C.3 enabled preZoRLA2 to avoid pattern instabilities while also being adaptive.

Results of preZoRLA2 operation

Operation of the Figure C.3 learning mechanism can lead to different patterns of functional zonation when used by 20 SSs connected in series (Figure C.1C) using the RESPONSE SIGNAL generation and detection mechanism illustrated in Figure C.1B. The patterns produced are always stable. In Figure C.6, the administered COMPOUND is a TOXIN. The five example results in Figure C.6 used the same parameterization, but the three TOXINS had different potencies (k_t values). The duration of repetitive exposure to the same TOXIN dose was long enough to produce stable patterns. Values of repeat experiments are not identical, but the patterns are identical. Prior to TOXIN dosing, each SS's *Q* value was initialized to zero. There are two measures of zonation. One is the

value of p_{AVG} (it maps to intrinsic clearance), which is the probability that a SS will clear a TOXIN if it is detected, averaged over the past N_{avg} = 500 simulation cycles. The second is TOXIN Elimination Count; it is the number of TOXINS CLEARED by a SS. In preZoRLA2, TOXIN Elimination Count maps to hepatocyte exposure. Consequently, if the TOXIN is HEPATOTOXIC, then we can expect measures of HEPATOTOXICITY to correlate with TOXIN Elimination Count. Note that when potency changes, each SS's Q value also changes, and that causes its p_{AVG} and TOXIN Elimination Count to change.

Figure C.6 shows examples of zonal patterns of clearance effort produced by preZoRLA2 for five different k_t values. k_t (the number of R-SIGNAL objects produced per exited TOXINS) controls the potency of the TOXIN. The average clearance efforts, p_{AVG} (blue bars), as well as TOXIN Elimination Count (red bars) by each SS are shown in Figure C.6A, C, E, G, I. Corresponding g and Q values are also shown for each pattern (Figure C.6B, D, F, H, J). When k_t = 0, only perivenous SSs expended clearance effort (Figure C.6A). As k_t increased, the response expanded toward the periportal region. The peak TOXIN Elimination Count moves toward the periportal region as well.

C.2 The 2D ZoRLA

Algorithm

The general sketch of the algorithm for the 2D ZoRLA is described below:

Initialize the (m by m) 2D space as shown in Figure 3.13a
Initialize Q and g (as described below)

Simulation cycle starts
For all SSs do {
   nSeenToxinsi ← 0
   nEliminatedToxinsi ← 0
   nExitedToxins ← 0
}
Randomly place toxDose toxins in the portal vein space.
While at least one toxin exists in space do {

- Move toxins (1 step) according the biased random walk algorithm.
- SSs see a co-located toxin with probability 1-pBypass; in that case update a counter: nSeenToxinsi ← nSeenToxinsi + 1
- SSs that see toxins randomly choose action according to their strategy p.
- For SSs who choose to eliminate do: nEliminatedToxinsi ← nEliminatedToxinsi + 1

}
nExitedToxins ← Number of toxins that reached the central region

For all SSs do **{**
   $\text{penalty}_i \leftarrow -(k_c \cdot \text{nEliminatedToxins}_i) - (k_a \cdot \text{nExitedToxins})$
}
Simulation cycle ends

Iterate *Q*-learning algorithm:
For all SSs do **{**

   Update Q_i nSeenToxin_i times:
     $Q_i \leftarrow (1-\alpha)Q_i - \alpha(\text{penalty}_i - \beta V)$
     where $V = \min(Q_i, g_i)$.

   Update p_i:
   $p_i = \exp(-g_i/h)/\{\exp(-g_i/h) + \exp(-Q_i/h)\}$
}
Start a new simulation cycle.

Legend:
SS: Sinusoidal Segment
$nSeenToxins_i$: number of TOXINS seen by SS_i.
$nEliminatedToxins_i$: number of TOXINS eliminated by SS_i
nExitedToxins: number of exited TOXINS at each cycle.
k_a: Penalty of detecting an RESPONSE signal object
k_c: Cost of eliminating a TOXIN
toxDose: Total number of TOXIN objects to be dosed
pBypass: Probability that a TOXIN bypasses a co-located SS

Derivation of Equation (3.7)

Q-learning is a form of reinforcement learning in which an agent attempts an action at a particular state, and depending on its consequences, receives reward or penalty. By trying all actions in all states over and over again, it learns the best strategy for choosing actions at a given state. In the original form of Q-learning algorithm (76), an agent at time t:

- observes its current state x_t,

- selects and performs an action at,

- observes the subsequent state y_t,

- receives a payoff *r*, and

- adjusts its Q values for xt and at according to:

$$Q_{t+1}(x, a) = (1 - \alpha)Q_t + \alpha(r + \beta V(y_t)) \quad \text{(A. 1)}$$

where $V(y_t) = \max_{a'}\{Q_t(y, a')\}$.

The probability of executing action a in state *x* is commonly determined using Boltzman equation (119):

$$p(a) = \frac{\exp(Q(x,a)/T)}{\sum_{a'}\exp(Q(x,a')/T)} \quad \text{(A. 2)}$$

In ZoRLA each agent has one state. There are two actions available to each, namely a_1 = *eliminate* and a_2 = *ignore*. Consequently, there are two Q values for each agent. Call them $Q(a_1)$ and $Q(a_2)$. Equation (A. 1) becomes:

$$Q_{t+1}(a) = (1 - \alpha)Q_t(a) + \alpha(r + \beta V) \quad \text{(A. 3)}$$

In this work r is always negative (penalty associated with detection of R-SIGNALS and toxicity costs). Thus Q values become negative after a few iterations. To make Q values positive and hence easier to interpret we specified the following variable changes:

$g = -Q(a_1)$

$Q = -Q(a_2)$

$h = T$

Substituting in equations (A. 1) and (A. 2) yields:

$$Q_{t+1} = (1 - \alpha)Q_t - \alpha(r - \beta V) \quad \text{(A. 4)}$$

where $V = \min\{Q_t, g\}$, and

$p_t = \exp(-g/h)/\{\exp(-g/h) + \exp(-Q_t/h)\}$

As described in Section 2.2, we specified g as a fixed noisy function of the SSs distance to the center of the grid.

Initializing Q and g

SSs determined their g values based on locals of some factor. We specified that downstream agents encounter significantly different factor levels than do upstream agents. Consequently, SSs' g values correlate with their distance from the exit area. Because

the exit area is at the center of the grid space we specified g as follows:

$g_i = (X_i^2 + Y_i^2) + \eta$, where X_i and Y_i are coordinates of the i^{th} SS with respect to the center of the grid space, and η is uniform random noise in the interval [0, m/2] (m is the length and width of the space). Q values were initialized such that agents started with uniform, randomly distributed clearance efforts. $Q = g + h \cdot ln\,(1/\text{PRN} - 1)$, where PRN is a uniformly distributed random number between 0 and 1.

Biased Random Walk

Particles (TOXINS) stochastically move towards the center of the grid (Figure C.4). For any given grid space, a flow direction, θ, is specified. Moor neighbors are numbered counter clock wise, starting with number 1 at zero degrees and ending with number 9 at 360° (Figure C.4A). In order to determine the stochastic movement of particles we specified a Normally distributed random variable (x), whose mean, $1 \leq \mu < 9$, corresponds to the effective flow direction, θ (for example $\mu = 4$ corresponds to $\theta = 135°$ and $\mu = 1.5$ to $\theta = 22.5°$). Details follow: A particle located at a grid space remains in place with probability p_{stop}; or stochastically moves to an adjacent space with probability $1 - p_{stop}$. The adjacent space, *nextSpace*, is stochastically chosen according to the flow characteristics as: $nextSpace = \lfloor x - 0.5 \rfloor \text{(modulo 8)} + 1$, where x is a Normally distributed random number: $x \sim N(\lfloor \theta/45 \rfloor + 1, \sigma^2)$, where $0° \leq \theta < 360°$ is the effective angle of movement (as shown in Figure C.4A) and σ is the standard deviation of x. σ controls the lateral movement of particles. Distributions are shown for three values of σ in Figure C.4B; the motion converges to Brownian motion for large values of σ (since the distribution becomes uniform).

In this work we used three different standard deviations: σ_1, σ_2 and σ_3, for periportal, middle and perivenous zones, respectively. Because the sinusoid density and therefore the lateral movement is greater periportal, we arbitrarily set $\sigma_1 = 1.2$, $\sigma_2 = 1.0$ and $\sigma_3 = 0.7$ (Figure C.4A).

Updating Q n times

Updating Q by Eq. (A. 5) (below) is equivalent to updating Q n times using Eq. (A. 4). Doing so reduces the computational complexity of the algorithm.

$$Q_{t+n} = a^n Q_t - (1 - a^n)(r - \beta V) \qquad \text{(A. 5)}$$

where $a = (1 - \alpha)$.

Proof:

$Q_{t+1} = (1 - \alpha)Q_t + \alpha(r - \beta V)$

$Q_{t+n} = (1 - \alpha)^n Q_t + \alpha\{\sum_{k=0}^{n-1} (1 - \alpha)^k\}(r - \beta V)$

$S = \alpha\{\sum_{k=0}^{n-1} (1 - \alpha)^k\} = \alpha\{(1 - \alpha)^{(n-1)} + (1 - \alpha)^{(n-2)} + ... + (1 - \alpha) + 1\}$ is a geometric series:

$S = 1 - (1 - \alpha)^n$

Thus:

$Q_{t+n} = (1 - \alpha)^n Q_t + \{1 - (1 - \alpha)^n\}(r - \beta V)$.

C.3 Switch-Like behavior of ZoRLA

Some CYP enzymes (such as CYP1A1 and CYP1A2) exhibit switch-like expression

behaviors when induced by TCDD (107, 108, 109), i.e. hepatocytes appear as either uninduced or fully induced. preZoRLA2 and ZorLA SSs can exhibit the same behavior when parameter h is small. Proof: expression $p = \exp{-(g/h)}/\{\exp(-g/h) + \exp(-Q/h))\}$ can be written as

$p = 1/\{1 + \exp(-(Q - g)/h)\}$. When $h \rightarrow 0$, if $(Q - g) > 0$, then $\exp\{-(Q - g)/h\} \rightarrow \infty$, and $p \rightarrow 0$; on the other hand if $(Q - g) < 0$, then $\exp\{-(Q - g)/h\} \rightarrow 0$, and $p \rightarrow 1$. In the rare case of $Q = g$, $p = 0.5$ regardless of h. Figure C.7 shows examples of such performance with $h = 0.01$, and dose = 50, 100, 500 and 1,000 TOXINS.

C.4 Single Dose Response

Can SSs in ZoRLA respond and adjust their clearance strategies following a single dose? We conducted experiments to answer that question. The results are shown in Figure C.8 and Figure C.9. SSs were pretreated with a COMPOUND (dose = 50 TOXINS) having zero potency for 200 cycles. This pretreatment was necessary to obtain the initial uninduced SS response similar to that of an uninduced liver. Given this initial condition, two types of dosing experiments were conducted: 1) three different doses of the same TOXIN (Figure C.8) and 2) identical doses of five different TOXINS having different potencies (Figure C.9).

SSs used the same algorithm as in Fig 6, except that extrahepatic RESPONSE signals were not generated. Instead, the SSs treated TOXINS as RESPONSE signals.

C.5 Patterns of Hepatotoxicity: Consequences of Changing k_a and *pBypass*

Figure C.10 shows various patterns of enzyme expression and hepatotoxicity in the

liver. Figure C.11 shows the effect of changing k_a and *pBypass* on patterns of SS damage.

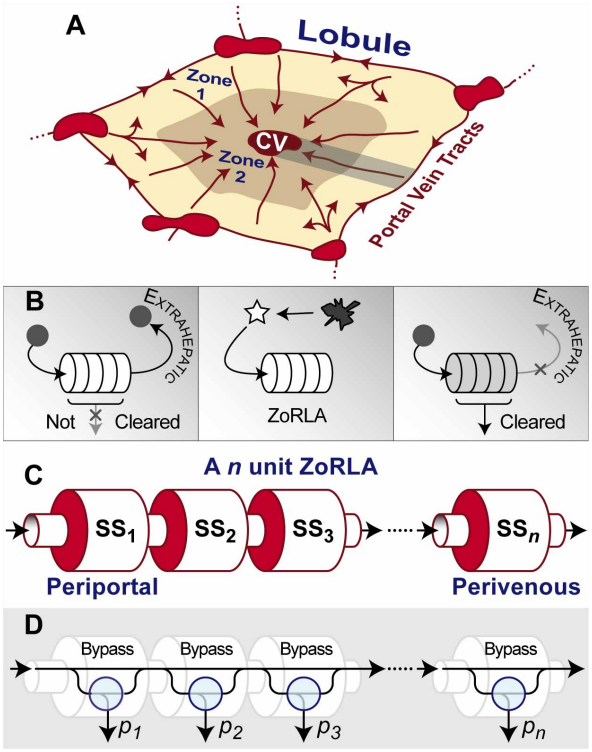

Figure C.1**.** Key features of the *in silico* experimental systems. (**A**) Illustrated is a hepatic lobule cross-section divided into two zones, periportal (Zone 1) and perivenous (Zone 2). (**B**) Illustrated is the TOXIN hepatic clearance scenario described in the text. Left: we start with a ZoRLA that has a low probability of clearing a TOXIN. A TOXIN enters and exits into the EXTRAHEPATIC space without being cleared. Center: The EXTRAHEPATIC TOXIN causes damage, which releases a RESPONSE SIGNAL object. The RESPONSE SIGNAL moves to and through the ZoRLA. In response to RESPONSE SIGNALS, mechanisms within each SS respond so that the probability of TOXIN clearance is increased. That adjustment is illustrated in the right panel by the change in SS shading. Right: with TOXIN clearance probability increased, the next TOXIN is more likely to be cleared: the risk of EXTRAHEPATIC TISSUE damage is reduced. (**C**) Illustrated is a ZoRLA composed of *n* quasi-autonomous SS. It maps to a periportal-to-perivenous portion of a lobule, such as the rectangular shaded region in **A**. Each SS has the same internal mechanisms. **(D)** Illustrated are low-resolution views of mechanisms within each SS agent. A TOXIN that enters a SS will bypass that SS (pass through undetected by the SS mechanisms) with a probability $pBypass_i$. The large circle in each SS_i represents its mechanisms. One mechanism is described in Figure C.2, and the other in Figure C.3. p_i, is the clearance strategy for SS_i; it is the probability that SS_i will clear a detected TOXIN.

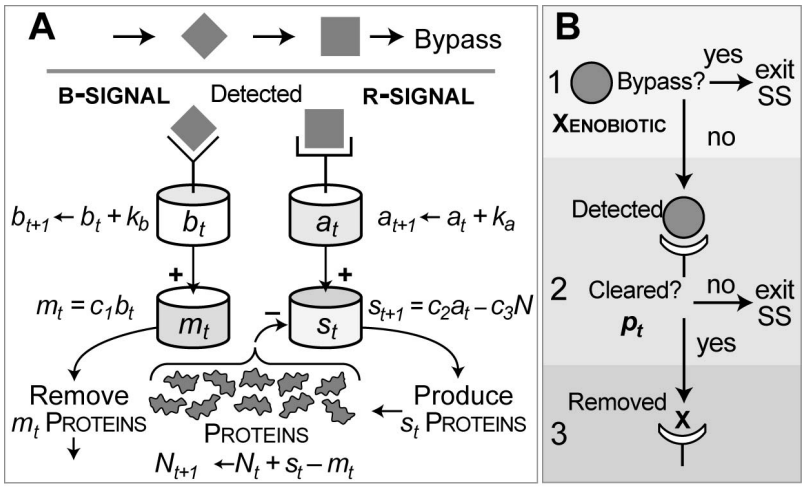

Figure C.2**.** A simple SS mechanism to create functional zonation across multi-SS ZoRLA. The events shown occur within each individual SS during each simulation cycle. See text for details. Mobile objects move from left (PERIPORTAL) to right (PERIVENOUS). At the start of each simulation cycle, all mobile objects (R-SIGNAL, B-SIGNAL, and XENOBIOTIC if present) move to the right (in Figure C.1) by one SS and new objects (one of each type) enter. (**A**) Each simulation cycle, the mechanism will determine if each mobile object present will or will not exit and visit the next, downstream SS. That process creates mobile object gradients along the ZoRLA's length. Subscript *t* indexes the value for the current simulation cycle. Subscript *t*+1 indexes the value for the next simulation cycle. When the *b* subsystem detects a B-SIGNAL, it increases b_t and that causes the removal of m_t PROTEINS and possible elimination of the B-SIGNAL. When the *a* subsystem detects an R-SIGNAL, it increases a_t and causes the production of s_t PROTEINS coupled with the illustrated negative feedback on the *a* subsystem and possible elimination of the R-SIGNAL. (**B**) A mechanism for XENOBIOTIC removal by each SS is illustrated. 1) A XENOBIOTIC that enters either bypasses that SS (and moves to the next SS) or is detected. 2) The probability of clearance following detection is p_t. 3) If cleared, it is removed from the simulation. p_t is a function of the number of PROTEINS *N*, analogous to xenobiotic clearance being a function of hepatocyte P450 levels.

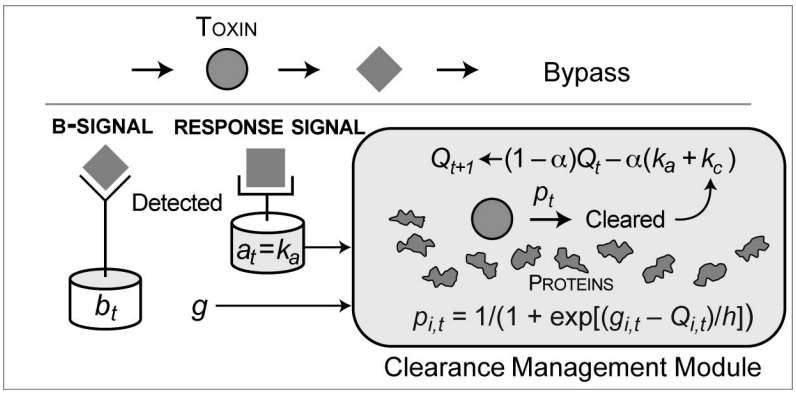

Figure C.3**.** A mechanism that enables each preZoRLA2 and ZoRLA SS to learn and adjust its TOXIN clearance strategy. preZoRLA2's stricture is as described in Figure C.1. The process within each SS is based on that in Figure C.2 and follows the scenario in Figure C.1B. The events shown occur within each individual SS during each simulation cycle (without knowledge of what other SS are doing). See text for details. The CLEARANCE-TOXIN management module (CT-module hereafter) is a quasi-autonomous SS subsystem (indicated by the PROTEIN objects) that manages TOXIN clearance. There is no cost during the current simulation for letting a TOXIN bypass to the next SS. RESPONSE SIGNALS do not pass undetected, but a TOXIN can bypass a SS with a probability specified by *pBypass*. A TOXIN, once detected, is cleared (eliminated) with probability p_t, which is current, updated clearance strategy of that SS. p_t is updated using the indicated equation. The value of Q_t (described in the text) is carried-forward from the preceding cycle. h is an adjustable parameter. g has a unique value for each SS, and, as described in the text; it provides information about relative SS location. The equation used to calculate the updated p_t value has the form of a Boltzmann distribution. When a RESPONSE SIGNAL was detected, the cost for each event is k_a. When a clearance event occurred, the cost of each is event k_c. Q_t is updated at the end of the simulation cycle for use during the next cycle using the indicated Q_t equation; it has the form of the well-known *Q*-learning algorithm.

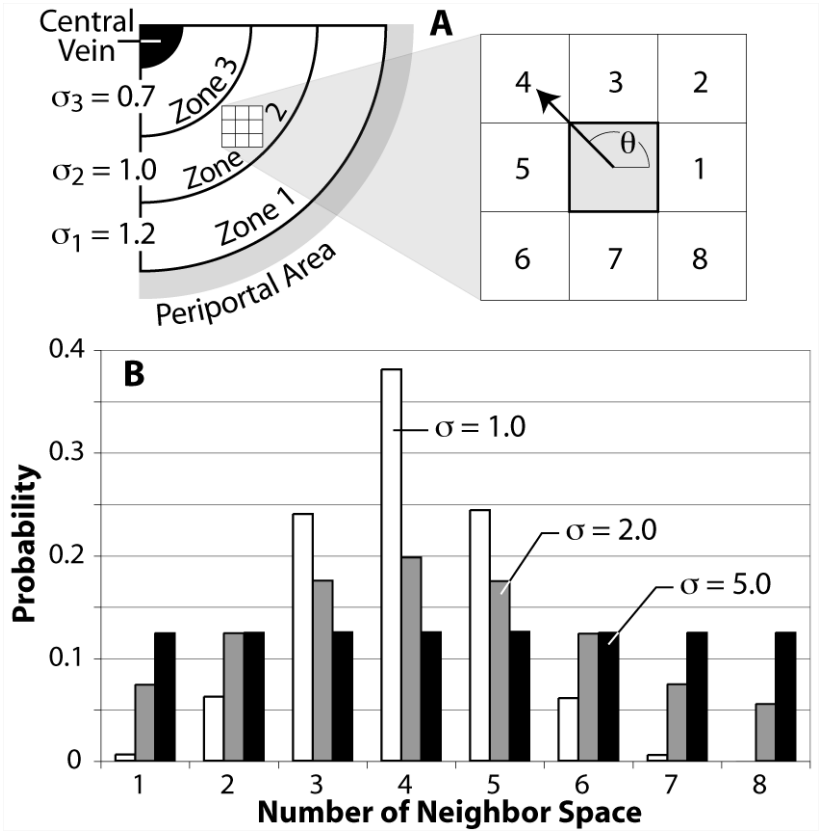

Figure C.4**.** Biased random walk algorithm. Toxin objects move stochastically on the grid space based on the biased random walk method described in section 1.3. (**A**) Flow direction is from the PERIPORTAL area (PV) towards the CENTRAL VEIN (CV). In order to simulate different upstream and downstream sinusoidal networks properties, the space is divided into three zones. Movement in Zone 1 has larger stochastic variance ($\sigma_1 = 1.2$) than does movement in Zone 3 ($\sigma_3 = 0.7$). A TOXIN's grid space (center, gray) and its eight neighbors are shown. The TOXIN moves to one of the eight neighboring spaces according to a semi-normal distribution with mean θ, and standard distribution σ. For the particular grid-space shown $\theta = 135°$. (**B**) Shown are three examples of the semi-normal distribution with mean = 4 (which corresponds to $\theta = 135°$) and standard deviations of 5, 2 and 1.

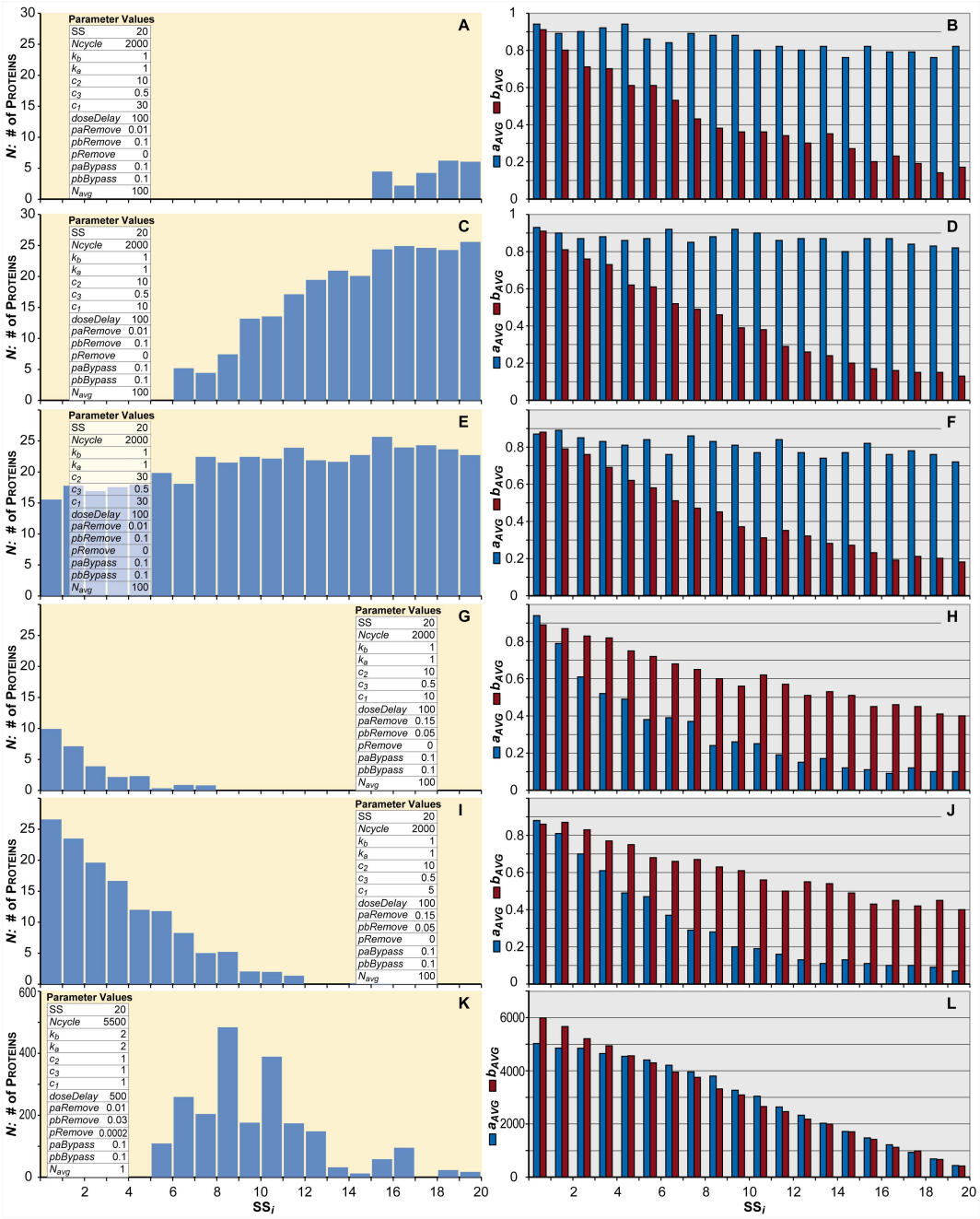

Figure C.5**.** Six examples of zonal patterns of PROTEIN expression by preZoRLA1. The simple mechanism of PreZoRLA1 is capable of generating various expression patterns: PERIVENOUS expression (**A**, **C**, **E**), PERIPORTAL expression (**G**, **I**), and mid-zonal expression (K) are shown for 20 SSs after an infusion (one COMPOUND per simulation cycle) lasting 2,000 simulation cycles. The infusion started after a delay of 100 cycles Average gradients of a-signal (a_{AVG}, blue) and b-signal (b_{AVG}, red) are shown for each expression pattern (**B**, **D**, **F**, **H**, **J**, **L**). The mechanism is shown in Figure C.2A. Parameters k_b, k_a, c_1, c_2, c_3, *doseDelay*, *paRemove*, *pbRemove*, *paBypass*, *pRemove*, *and* N_{avg} are described in Figure C.2 and the main text.

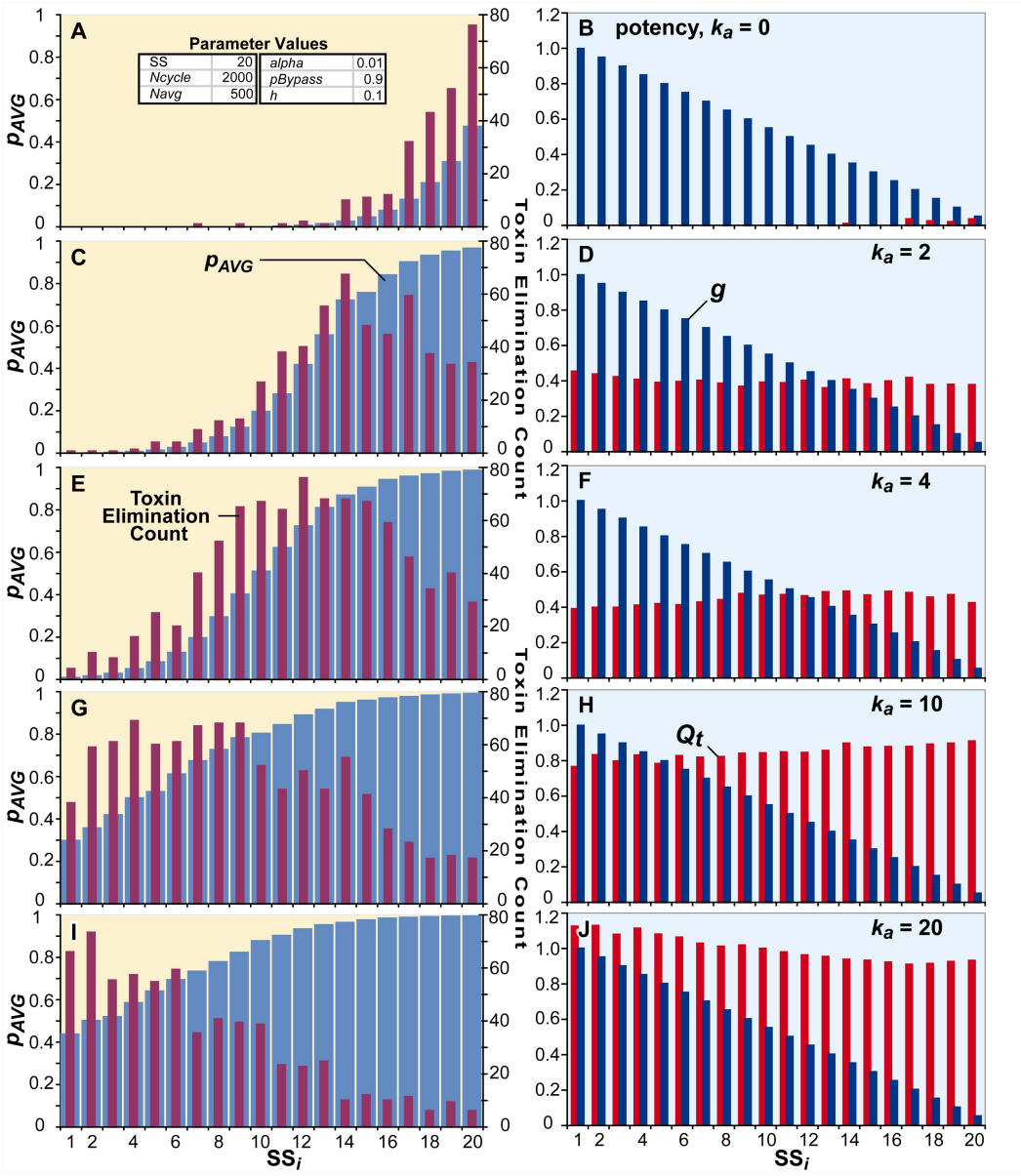

Figure C.6**.** Examples of zonal patterns of clearance effort by preZoRLA2 for five TOXINS having different potencies are shown. The zonation patterns are created by 20 SSs in series, each running the learning enabled, TOXIN clearance mechanism in Figure C.3. The average clearance efforts, p_{AVG} (blue bars), as well as TOXIN Elimination Count (red bars) by each SS are shown (**A**, **C**, **E**, **G**, **I**). p_{AVG} is the probability that a SS will clear a TOXIN if it is detected, averaged over the past N_{avg} = 500 simulation cycles. TOXIN Elimination Count is the number of TOXINS cleared by SS_i. Corresponding g and Q_t values are also shown for each pattern (**B**, **D**, **F**, **H**, **J**). (**A** and **B**) k_a (toxin potency is the same as response potency) = 0; (**C** and **D**) k_a = 2; (**E** and **F**) k_a = 4; (**G** and **H**) k_a = 10; (**I** and **J**) k_a = 20. The apparent oscillations in TOXIN elimination count are a consequence of multiple random events; they begin to vanish as values from several simulations are averaged.

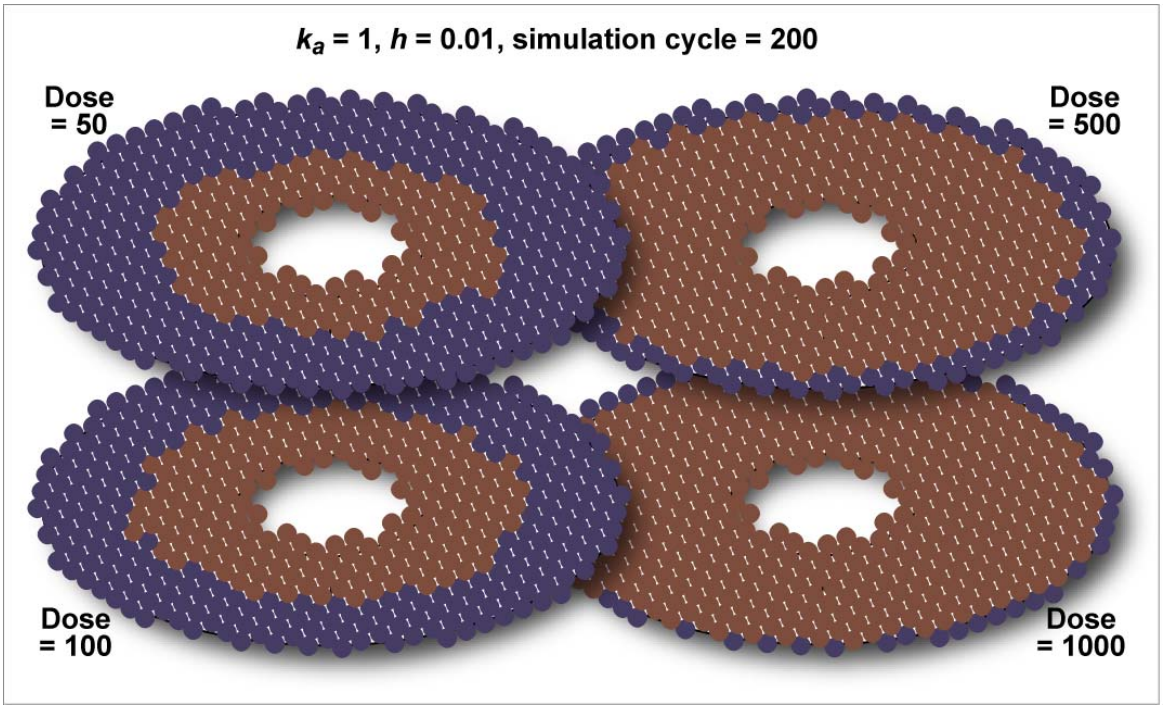

Figure C.7**.** Patterns of switch-like behavior. With *h* having a small value ZoRLA exhibits switch-like clearance behavior. The ZoRLA mechanisms and experimental protocol are the same as in text Figure 3.12. The color scale (shown below) too is the same. Four experiments were conducted. Each simulation cycle began with the indicated dose (50 to 1,000 TOXINS for k_a = 1). Shown is the p_i pattern at the conclusion of the 200th cycle. h = 0.01.

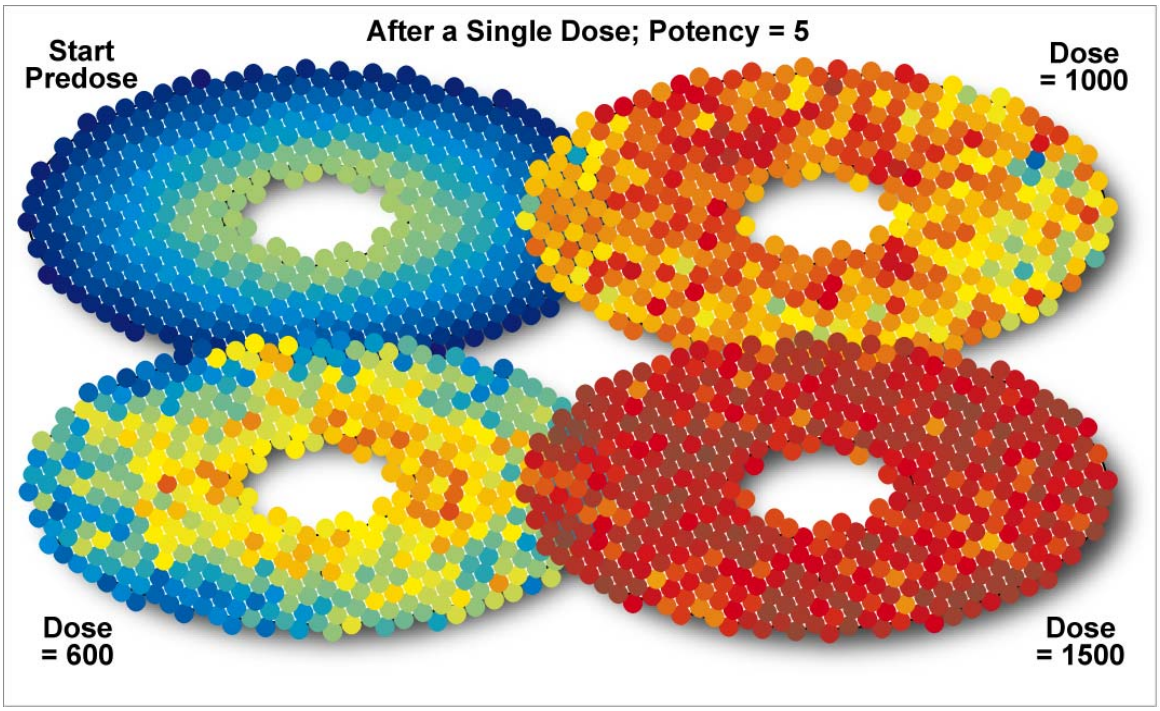

Figure C.8**.** Patterns of single-dose response. The ZoRLA mechanisms and experimental protocol are the same as in text Figure 3.13. Except that extrahepatic RESPONSE signals were not generated. Instead, TOXINS were used as RESPONSE signals. The color scale (shown below) too is the same as in text Figure 3.13. Three experiments were conducted. SSs were pretreated with a COMPOUND (DOSE = 50) having zero potency for 200 cycles to obtain the *Start Predose* condition shown. Each simulation cycle began with the indicated dose (600, 1,000 and 1,500 TOXINS for k_a = 5). Shown are the p_i patterns after one simulation cycle.

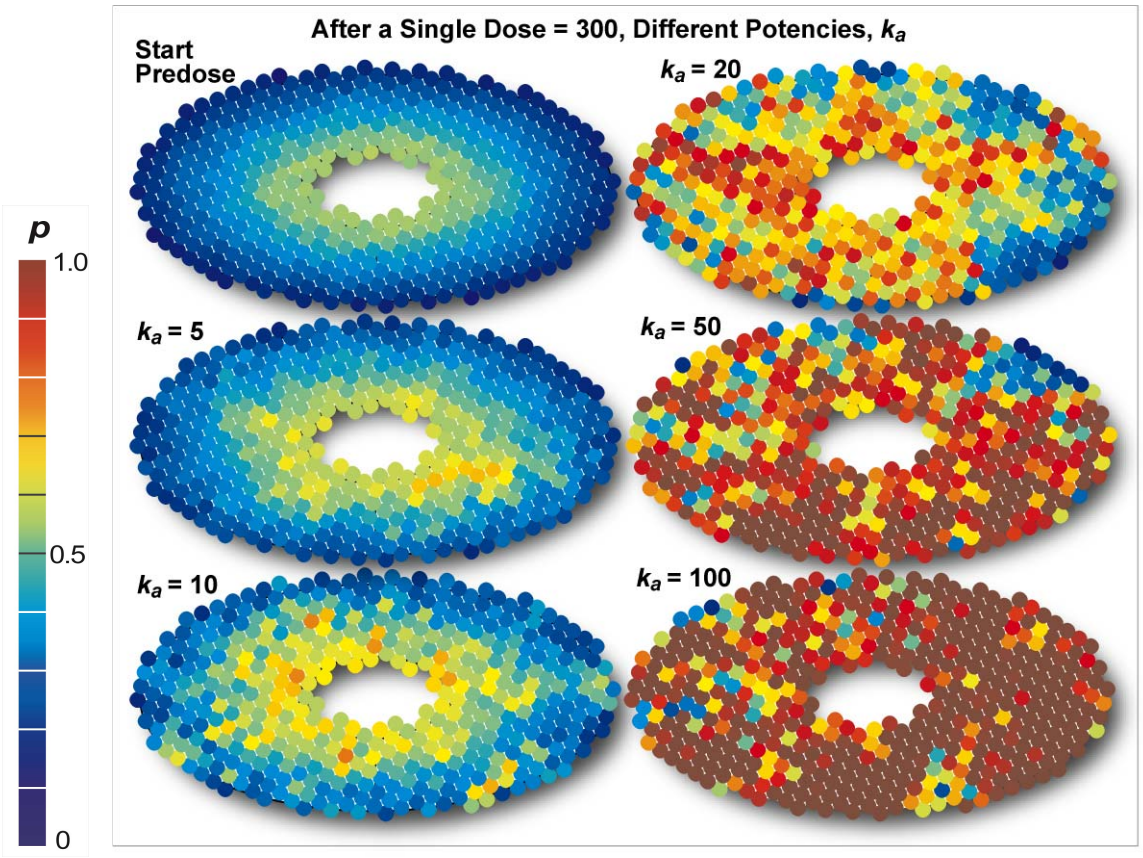

Figure C.9**.** Additional patterns of single-dose response. The ZoRLA mechanisms, experimental protocol and the color scale are the same as in text Figure 3.13, except that extrahepatic RESPONSE signals were not generated. Instead, the SSs treated COMPOUNDS as RESPONSE signals. Five experiments were conducted. SSs were pretreated with a COMPOUND (DOSE = 50) having zero potency for 200 cycles to acquire the *Start Predose* condition shown. Each simulation cycle began with the indicated potency (k_a = 5, 10, 20, 50 and 100) and DOSE = 300. Shown are the p_i patterns following one simulation cycle after the *Start Predose* condition.

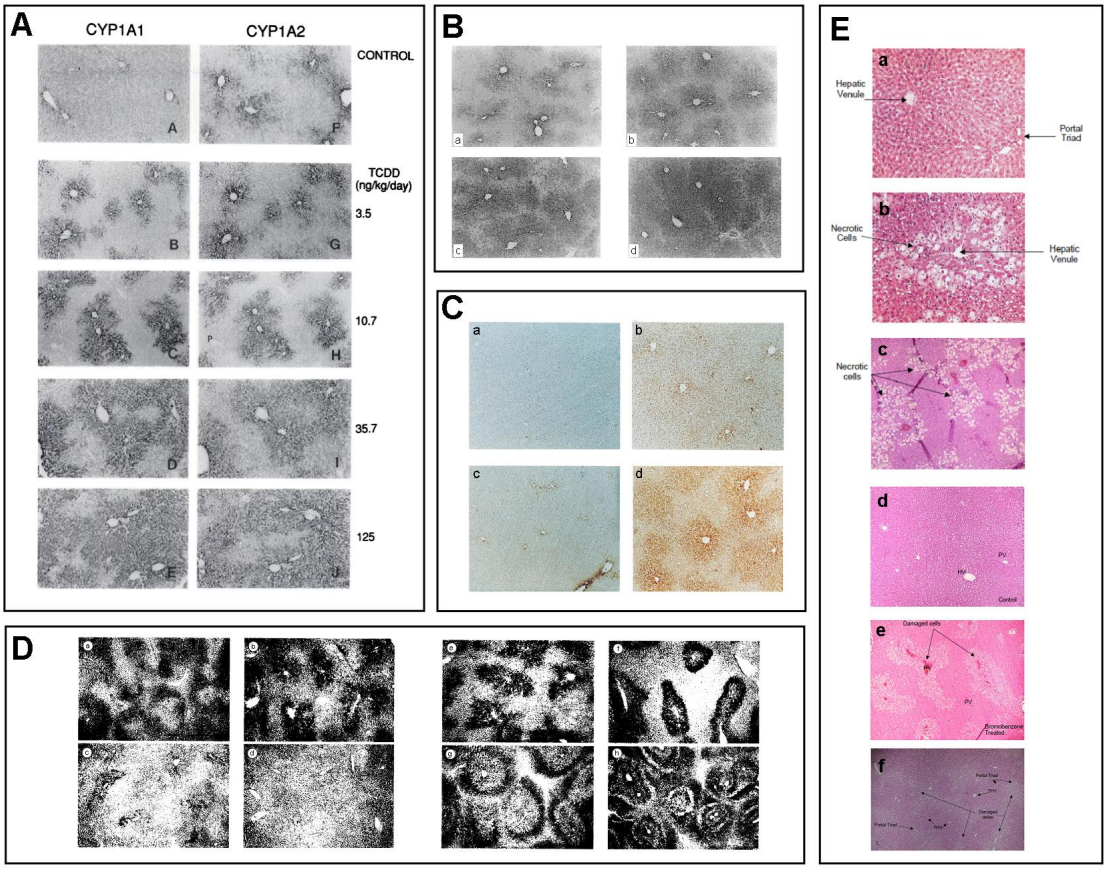

Figure C.10**.** Various patterns of enzyme expression and hepatotoxicity in the liver. (**A**) Induction of CYP1A1 & CYP1A2 expression by TCDD in rats. TCDD was administered for 30 weeks. Daily dose amounts are shown on each image (110). (**B**) Induction of CYP2B1/2 expression by D4 in rats. D4 was administered for a period of 5 consecutive days 6 hours/day. Dose amounts are (**a**) control, (**b**) 30 ppm, (**c**) 300ppm, (**d**) 700 ppm (111). (**C**) Induction of CYP1A1 (**a**, **b**) & CYP1B1 (**c**, **d**) expression by TCDD in rats. TCDD was administered for a period of 30 weeks biweekly at the following daily dose values (**a**, **c**) control, (**b**, **d**) 35 ng/kg (112). (**D**) Hepatotoxicity (lipid peroxidation and cell necrosis) patterns by Bromotrichloromethane ($CBrCl_3$) perfusion. Damaged cells are shown in black. Dose amounts and infusion durations are as follows (**a**) 15 min $CBrCl_3$, (**b**) 30 min $CBrCl_3$, (**c**) 60 min $CBrCl_3$, (**d**) 30 min $CBrCl_3$ & N,N'-diphenyl p-phenylene diaminel, (**e**) 30 min $CBrCl_3$, (**f**) 15 min $CBrCl_3$ (20% O_2), (**g**) 60 min $CBrCl_3$ (20% O_2) trypan blue uptake, (**h**) 60 min $CBrCl_3$ (20% O_2) co-staining with guchsin (115). (**E**, **a**-**c**) Perivenous hepatotoxicity (cell necrosis) patterns by carbon tetrachloride (CCl_4). A single dose (1.0 mmol/100 g body weight) of CCl4 was administered following a 24-hour fast. Damaged cells are shown in white. (**E**, **d**-**e**) Perivenous hepatotoxicity patterns by bromobenzene. A single dose (3.8 mmol/kg body weight) of bromobenzene was administered following a 24-hour fast. Damaged cells are shown in white. (**E**, **f**) Periportal hepatotoxicity patterns by antegrade digitonin perfusion (5.0 mg/ml) for 90 seconds. Damaged cells are shown in white (116)

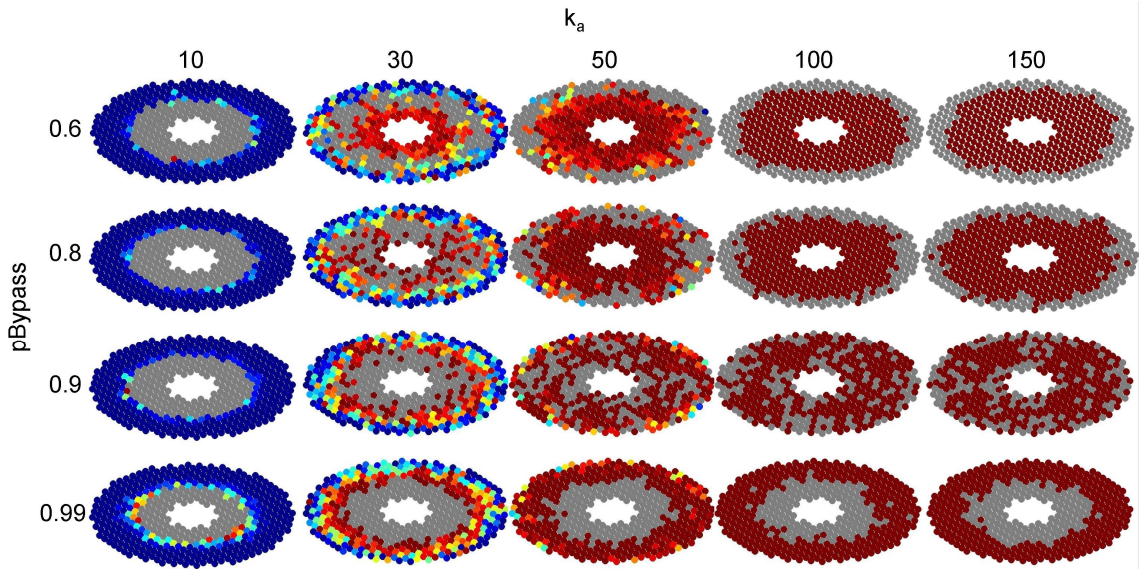

Figure C.11. Effect of changing k_a and *pBypass* on patterns of SS damage. The ZoRLA, experimental protocol, and *p* color scale are the same as in Figure C.9. Damage (which can map to either severe hepatocyte damage or necrosis) occurs after any SS_i cleared 50 COMPOUNDS. A damaged SS is gray. Simulations were terminated when one-third (152 out of 460) of SSs died. SSs were pretreated with a COMPOUND (DOSE = 50, *h* = 70, *pBypass* = 0.8) having zero potency for 200 cycles to acquire the *Start Predose* condition shown in Figure C.9. The DOSE each simulation cycle was 50 COMPOUNDS. For the patterns shown *h* = 10. Using the default value *h* = 70 used for article Figures 3.15–3.18, we were unable to clearly demonstrate a diverse range of toxicity patterns for the selected ranges in *pBypass* and k_a values. We observed that diversity increased for smaller *h* values. We selected *h* = 10 because it provided and interesting variety of patterns.

Appendix D. Response to Reviewers

Chapter 3 was originally adapted from a manuscript submitted to the Journal of Theoretical Biology for publication. The manuscript was later revised to address the journal reviewers' concerns. A summary of major revisions made to the manuscript along with the response to important reviewers' questions are listed below.

Reviewer 1

Reviewer 1: One of the strengths of the presented method is the reliance upon experimentally identified behaviors; however it is unclear as to how much prior experimental data needs to be obtained, and this raises the question of the specific added value of the resultant model? If different compound properties result in different zonation patterns, then the data for those properties needs to be generated prior to modeling: if this is the case, then what is gained by the model?

Reply: We posit that a specific mechanism may be responsible for enzyme induction within each hepatocyte. Differences in compound properties result in relative differences in the degree to which the compound interacts with mechanism components. Our hypothesis is that differences in exposure of a given hepatocyte as a consequence of its relative location within the lobule, in combination with other factors, is sufficient to cause dose, time, and zone dependent patterns in the levels of metabolizing enzyme. Pattern similarities between *in silico* and wet-lab observations provide a degree of validation for the implemented mechanism; more importantly, however, if the *in silico* model generates patterns different from those generated by the wet-lab model, then the

micromechanisms upon which the *in silico* model are based are likely incorrect. That is a primary benefit gained only by modeling.

Another value of the model is in having a concrete, validated mechanism that can be challenged. A further value is being able to predict spatially heterogeneous induction (and possibly toxicity) patterns in whole livers given simple data obtained *in vitro*. If the relative level of interaction of a new compound can be measured in a simple *in vitro* test, then that information may be sufficient to parameterize a future ZoRLA and anticipate what dose, time, and zone dependent patterns that compound may generate.

To make these points clear, we added the following as the penultimate paragraph under Discussion.

The 2D ZoRLA provides value in three ways. 1) When a specific ZoRLA mechanism generates patterns different from those observed in wet-lab data, then the micromechanisms used in the model are likely incorrect and in need of modification. 2) Having achieved a degree of validation, the model's concrete micromechanisms can be challenged experimentally. 3) We have demonstrated the feasibility of predicting spatially heterogeneous induction (and toxicity) patterns in whole lobules for a new compound, given a validated ZoRLA and data that can be obtained *in vitro*. If the relative level of interaction of a new compound can be measured in a simple *in vitro* test, then that information may be sufficient to parameterize the ZoRLA and anticipate what dose, time, and zone dependent patterns that compound may generate and if toxicity is likely.

Reviewer 1: On a related point, in the Discussion, the authors comment on not increasing model granularity prematurely, and then follow with a discussion of the

process of addressing the inverse mapping problem (phenomenon to generators). I agree with the comments, however, am interested in 1) how the authors would suggest the mapping from proposed generators to identified signaling/metabolic pathways. 2) Do these pathways need to be described in terms of correspondingly mapped functions? 3) How can those functions be parsed experimentally from the overall behavior of the cells (as they act in a traditional lab context)? This process of mapping is necessary if the multiple plausible generators are to be evaluated.

1) The current abstract micromechanisms map to (a conflation of) all fine-grained processes in the referent that contribute to the simulated event. If we do not have specific evidence on how some known signaling &/or metabolic pathway is contributing, then there is no scientific value in simply implementing some conceptual linkage within a ZoRLA simply for the sake of including it.

Consider the following. We have observations from identical experiments on two different livers, one normal (wild type) and another from a mouse that has had one or more components within a signaling &/or metabolic pathway measurably altered, e.g., by genetic deletion. Wet-lab data from the knockout's liver (but not that from the normal liver) falsifies the current micromechanism. That evidence forces us to posit one or more new, more fine-grained micromechanisms that incorporate one or more features containing the knocked-out component.

Such an approach was used by Tang and Hunt (122) to falsify a coarse-grained micromechanism and replace it with one more fine-grained in which components mapped directly to individual macromolecules.

2) The question is answered above.

3) This question is partially answered above. An experiment, such as the above knockout experiment, can be proposed because its results are expected to falsify one or more competing, equally plausible mechanisms. An effective alternative strategy is to expand the variety of systemic attributes targeted. Addition of a new attribute can falsify a micromechanism. To revalidate, it may be necessary to replace a coarse-grained micromechanism with one that is somewhat more fine-grained. Following that, a clearer mapping may exist between the more fine-grained micromechanism and the signaling and/or metabolic pathways of interest. Further iterative refinement can increase the concreteness of the mapping. Such a process was used by Lam and Hunt (98) to move iteratively from coarser to more fine-grained micromechanistic hypotheses.

Part of the preceding text has been added to the Discussion Section 3.2.4.

Reviewer 2

Reviewer 2 (R2), point 1: No comparison or contrast with alternative modeling and simulation approaches is provided.

Because synthetic analogues (e.g., ZoRLAs) and the familiar inductive, mathematical liver and hepatic zonation models are intended for fundamentally different uses, they are not easily compared. We should have pointed that out. To address this point, we did the following. 1) We added a citation to an additional conceptual model (123). 2) We deleted the last half of the original 2nd paragraph of the Introduction and added a new paragraph. In it we draw attention to detailed model of Ohno et al., (120) and point the reader to the recent, extensive review by Ierapetritou et al., (121) of liver

models. That review focuses on zonation of xenobiotic-metabolizing enzymes. We also state, “The computational modeling and simulation (M&S) approach used herein (Figure 3.9) and the resulting models are fundamentally different from those inductive mathematical models and so are not directly comparable. Hunt et al. (2) explain those differences and how the two different M&S approaches complement each other.”

R2, point 2: It is important to define and motivate the models. To do so, all variables, constants and parameters should be defined contiguous with the statement of a model. For example, it is not sufficient to state c2 and c3 are tunable constants as done in the text in Supplementary Material without further description until they are used in Figure. C.2.

Reply: We added the following to section C.1 to more clearly describe the roles of c_1, c_2, & c_3.

“c_2 and c_3 are tunable constants, which adjust the influence of a_{AVG} and N_{avg} on s. Larger c_2 values increase PROTEIN production; c_3 modulates negative feedback on PROTEIN production by already existing PROTEINS, dampening the impact of a change in R-SIGNALS on PROTEIN levels.”

R2, point 3: In section 3.2.2, all symbols should be defined after Eq. 3.6 and Eq. 3.7 precisely as used in these equations. Elsewhere, perhaps in a table, they should be defined and their units and range of values specified.

Reply: We added the following tables.

Table D.1. Parameters of preZoRLA1.

Symbol	Description	Range	Used
N	Amount of PROTEINS expressed	[0, ∞)	*0 – 500
pbBypass	Probability that a B-SIGNAL bypasses an SS without being detected	[0,1]	0.1
paBypass	Probability that an R-SIGNAL bypasses an SS without being detected	[0,1]	0.1
pbRemove	Probability that a B-SIGNAL is removed following detection	[0,1]	0.03 – 0.1
paRemove	Probability that an R-SIGNAL is removed following detection	[0,1]	0.01 – 0.15
b_t	The value of *b*-subsystem at simulation cycle *t*: has an effect on down regulation of PROTEINS	[0, ∞)	*0 – 6000
a_t	The value of *a*-subsystem at simulation cycle *t*: has an effect on up regulation of PROTEINS	[0, ∞)	*0 – 6000
m_t	Amount of PROTEINS degraded during cycle *t*	[0, ∞)	*0 – 6000
s_t	Amount of PROTEINS expressed during cycle *t*	[0, ∞)	*0 – 5000
k_b	Increment of *b* following detection of a B-SIGNAL	[0, ∞)	1 – 2
k_a	Increment of *a* following detection of an R-SIGNAL	[0, ∞)	1 – 2
c_1	Proportionality constant which relates *b* to *m*	[0, ∞)	1 – 30
c_2	Constant that regulates positive influence of *a* on *s* (Figure C.2)	[0, ∞)	1 – 10
c_3	Constant that regulates negative influence of *N* on *s* (Figure C.2)	[0, ∞)	0.5 – 1

* Simulation result

Table D.2. Parameters of ZoRLA and preZoRLA2.

Symbol	Description	Range	Used
$p_{i,t}$	Clearance strategy of SS_i during simulation cycle *t*: the probability that SS_i clears a detected COMPOUND during simulation cycle *t*	[0,1]	*0 – 1
$Q_{i,t}$	Estimate of SS_i's long-term, discounted cost at simulation cycle *t*	(-∞,∞)	*0 – 3500 ZoRLA *0 – 1.2 preZoRLA2
g_i	The value of the P-to-P B-SIGNAL gradient used by SS_i	[0, ∞)	0 – 200 ZoRLA 0 – 1 preZoRLA2
pBypass	Probability that a TOXINS bypasses an SS without being detected	[0,1]	0.6 – 0.99 ZoRLA 0.9 preZoRLA2
h	A constant which adjusts the randomness of decisions	[0, ∞)	0.01 – 70 ZoRLA 0.1 preZoRLA2
α	Learning rate: determines the weight of newly observed costs in calculation of Q value	[0,1]	0.1

Symbol	Description	Range	Used
β	Discount factor: determines the current importance of future costs	[0,1]	0.5 ZoRLA 0 preZoRLA2
k_a	Cost of detecting one R-SIGNAL	[0, ∞)	0 – 150 ZoRLA 0 – 20 preZoRLA2
k_c	Cost of clearing one COMPOUND	[0, ∞)	1
	Aditional parameters for ZoRLA, only		
(X_i,Y_i)	Coordinates of SS_i with respect to the center of the grid space	$\Re^2$	(4,4) – (28,28)
η	Uniform random noise	[0, grid radius]	[0, 15.5]

* Simulation result

R2, point 4: Parameter h has the role of an excitation temperature in the search algorithm, and h = 0.01, 5, 10 and 70 is used in various simulation experiments. It would be valuable to provide some explanation and motivation for each choice (as done for h = 0.01).

Reply: The parameter *h* acts as a gain: it amplifies and/or dampens the difference: g – *Q*. A large value of *h* eliminates zonation completely; a small value magnifies it. In the revision we now state that in the paragraph following Eq. (3.6)

For a given condition and SS arrangement, values of *h* were explored to find those that, in combination with other parameter values, resulted in behaviors similar to those targeted. The default value was *h* = 70 for Figures 3.13–3.16 in the revised manuscript. However, using that value, we were unable to clearly demonstrate the compound elimination shift effect shown in the current Figure 3.12 for the selected range in ka values. We observed that the shift effect became more evident for smaller *h* values. We selected *h* = 5 to demonstrate that peak COMPOUND elimination count can shift from PERIVENOUS to PERIPORTAL as ka increases. We added this explanation to the Figure 3.12

legend.

Similarly, for the results in Figure C.11, to clearly demonstrate a range of toxicity patterns for the selected ranges of *pBypass* and *ka* we observed increasing diversity with decreasing *h* values. We selected *h* = 10 because it provided an interesting variety of patterns. We added this explanation to the Figure C.11 legend.

R2, point 5: In Figure 3.13, parameter g appears to vary as a function of ka. Why, has the discretization changed? If so, what is it for each value of ka?

Reply: g does not vary as a function of *ka*. To mimic biological variability, and as specified in the Supplement (Appendix C), a small amount or random noise is added to the sum of the squared coordinate values of the i^{th} SS each simulation. Consequently, the appearance of the graph of g changes slightly for each simulation.

R2, point 6: Figure C.6 shows oscillations in TOXIN Elimination Count. That is, after a loca maximum, there is a reduction in the SSi that follow. For instance in panel S6I, maxima at 2, 4, 6, 8, etc. are followed by a rapid persistent, decrease. Is this primarily a result of the discretization (the number of SS in the model)? Are oscillations observed in wet-lab experiments? Please elaborate.

Reply: These apparent oscillations are simply a consequence of multiple random events. As the experiments are repeated and results averaged, the apparent oscillations begin to vanish. We added a statement to that effect in the legend to Figure C.6.

R2, point 7: Results fit to experimental data in the literature (Figure 3.16) with equations chosen without providing any explanation for the choice of the form of these equations (other than they fit best) is not appropriate.

Reply: Because of the nature of synthetic modeling, producing results that fit the experimental data is not among intended model uses. The goal was building working mechanisms that can exhibit biomimetic phenomenon. Simply fitting the data, when abundant, can be easily done by means of inductive mathematical modeling. However, doing so does not concretize plausible underlying mechanisms.

To compare ZoRLA and wet-lab experimental results, a mapping method was required. Simple linear mapping was explored first, and that worked for the y-axis mappings. However, that failed for x-axis mappings. Consequently, log-linear mapping was explored, and it proved satisfactory. In Figure 3.16, two quantities were mapped:

1. P_{AVG} was mapped to “log Cyp1A2 mRNA expression”, and

2. COMPOUND elimination count (CEC) was mapped to “[TCDD] per viable hepatocyte”.

The former (y-axis) used a linear mapping for both parts b and c. The latter used log-linear mappings. Figure 3.16b and c used different x-axis mappings. A better match was achieved in part c. Note that the reason for the improvement was not that different mapping methods were used. It was because the ZoRLA parameterization changed.

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

Please sign the following statement:

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.

Author Signature

6/8/2010

Date