Predictions of Hepatic Disposition Properties Using a Mechanistically Realistic, Physiologically Based Model

Li Yan, Shahab Sheihk-Bahaei, Sunwoo Park, Glen E. P. Ropella, and

C. Anthony Hunt

The UCSF/UCB Joint Graduate Group in Bioengineering, University of California, Berkeley and San

Francisco, CA, USA: LY, SSB, CAH

Department of Bioengineering and Therapeutic Sciences, The BioSystems Group,

The University of California, San Francisco, CA, 94143-0446: SP, GEPR, CAH

Tempus Dictum, Inc., 27891 S.E. Starr Rd. Eagle Creek, OR 97022-8664: GEPR

1

Running title: Predictions of hepatic disposition properties

Correspondence: C. Anthony Hunt, Department of Bioengineering and Therapeutic Sciences, University

of California, 513 Parnassus Ave., S-926, San Francisco, CA 94143-0446

P: 415-476-2455; F: 415-514-2008   <a.hunt@ucsf.edu>

Number of text pages: 22

Number of figures: 7

Number of tables: 5

Number of references: 25

Words in Abstract: 240

Words in Introduction: 746

Words in Discussion: 1490

Abbreviations:

ANN: artificial neural network; CC: correlation coefficient; CV: central vein; FCMA: Fuzzy c-Means

algorithm; f_{uB}: unbound drug fraction; HBA: hydrogen bond acceptor; HBD: hydrogen bond donor; ISL:

In Silico Liver; PCP: physicochemical property; MW: molecular weight; P_{app}: octanol/water partition

coefficient for pH 7.4; PB: physiologically based; PK: pharmacokinetics; PV: parameter value; RB:

number of rotatable bonds; SD: standard deviation; SS: sinusoidal segment; TPSA: topological polar

surface area.

DMD #19067

Abstract

Quantitative mappings were established between drug physicochemical properties (PCPs) and parameter values of a physiologically based, mechanistically realistic, In Silico Liver (ISL). The ISL plugs together autonomous software objects that represent hepatic components at different scales and levels of detail. Microarchitectural features are represented separately from the mechanisms that influence drug metabolism. The same ISL has been validated against liver perfusion data for sucrose and four cationic drugs: antipyrine, atenolol, labetalol, and diltiazem. Parameters sensitive to drug-specific PCPs were tuned so that ISL outflow profiles from a single ISL matched in situ, perfused rat liver, outflow profiles of all five compounds. Quantitative relationships were then established between the four sets of drug PCPs and the corresponding four sets of PCP-sensitive, ISL parameter values; those relationships were used to predict PCP-sensitive, ISL parameter values for prazosin and propranolol given only their PCPs. Relationships were established using three different methods: 1) a simple linear correlation method, 2) the Fuzzy c-Means algorithm, and 3) a simple artificial neural network. Each relationship was used separately to predict ISL parameter values for prazosin and propranolol given their PCPs. Those values were applied in the ISL used earlier to predict the hepatic disposition details for each drug. Although we had available only sparse data, all predicted disposition profiles were judged reasonable (within a factor of two of referent profile data). The order of precision, based on a Similarity Measure, was $3 > 2 > 1$.

Introduction

The In Silico Liver (ISL) (Hunt et al., 2006) (Yan et al., 2007) (Fig. 1) is a first generation example of

synthetic, discrete, physiologically based, analogues that are intended for refining, exploring, and testing

hypotheses about the details of hepatic drug disposition. Autonomous components represent spatial

aspects of hepatic organization and function. Different drugs can be represented and studied

simultaneously or separately. Each ISL component can interact uniquely with any drug-representing

object that enters its local environment. The consequences of simulated systemic and local interactions

can be measured and studied simultaneously, analogous to how wet-lab experiments are conducted.

In (Yan et al., 2007), the simulated hepatic disposition of atenolol, antipyrine, labetalol, and diltiazem,

along with co-administered sucrose, used only one parameterized, ISL structure for all compounds. A

subset of components interacted differently with the particular compounds. Monte Carlo ISL variants

simulated compound-specific outflow profiles that matched referent profiles. The results supported two

hypotheses. 1) The mappings in Fig. 2 between ISL components and corresponding liver components

were sufficiently realistic for the stated model use. 2) The simulated drug-ISL-component interaction

events mapped to corresponding hepatic disposition events. A goal for this project has been to discover

and verify ISL counterparts to the relationships between drug physicochemical properties (PCPs) and

pharmacokinetic (PK) mechanisms, and use them to make predictions.

A vision motivating research on this class of models is identical to one that has motivated

development of traditional physiologically based (PB) PK models: by "accounting for the causal basis of

the observed data, … the possibility exists for efficient use of *limited drug-specific data* (italics added) in

order to make reasonably accurate predictions as to the pharmacokinetics of specific compounds, both

within and between species, as well as under a variety of conditions" (Rowland et al., 2004). PCP-

sensitive, PBPK model parameters necessarily conflate features and properties of the biology (aspects of

histology, etc.) with drug PCPs. In doing so, there is a risk that "the causal basis" becomes obscured due

to the conflated biological features that were especially influential in causing some property of the data.

Interconnections between sinusoids might be such a feature. In the ISL, because we have built a

collection of mechanisms from finer-grained components, we have precise control over conflation, yet the

causal basis is still there in the drug-component interaction logic (axioms, rules). Our expectation has

been that at some level of granularity, the complexity will be sufficiently unraveled so that the logic for a

given ISL component-drug interaction (the "causal basis") will rely heavily on only a small subset of

easily specified, drug and biological attributes. In some cases, information about the molecular attributes

may be represented adequately by just one or a few PCPs (or, more specifically, knowledge of the drug-

environment interaction for which a PCP is a measure). Once we have validated ISL models that exhibit

some of these characteristics, we will have moved closer to the above vision of being able to accurately

anticipate the hepatic disposition properties of a new compound, given only its molecular formula and

structure (and/or validated PCPs).

Are the mechanisms in the current ISL sufficiently realistic to enable such predictions using only

limited data? We hypothesized that the differences in the tuned values of the ten PCP-sensitive ISL

parameters were due primarily to differences in PCPs. If that is the case, then quantitative PCP-to-ISL

mappings exist and, when discovered, should enable us to predict the disposition of a new compound

given its PCPs. Failure to provide better than ballpark predictions could be taken as evidence invalidating

that hypothesis and possibly aspects of the ISL mechanisms. The goal of this project has been to suggest

quantitative PCP-to-ISL mappings and then use them to predict the disposition properties of prazosin and

propranolol given only their PCPs.

Three different methods were used to implement quantitative PCP-to-ISL mappings: linear

regression, an artificial neural network, and a well-established fuzzy clustering algorithm. Each method

predicted the PCP-sensitive, ISL parameter values for prazosin and propranolol. Those predicted values

were then combined with the already-validated, drug-insensitive ISL parameter values. Simulation of the

resulting ISLs gave three different, independent versions of expected hepatic disposition details for

prazosin and propranolol along with expected liver perfusion outflow profiles. Those profiles were

surprisingly good matches to the observed profiles, strengthening the ongoing validation of ISL

mechanisms and supporting the existence of the mapping relationships in Fig. 2. Together they represent

an important advance in our ability to predict PK properties for drugs given only their structure.

Methods

To distinguish clearly in silico compounds and processes from corresponding hepatic structures and

processes, we use SMALL CAPS when referring to the former.

Model Structure and Design. ISL structure is illustrated in Fig. 1. Brief summaries of its novel

structure and design follow for convenience. Consult (Hunt et al., 2006) and (Yan et al., 2007) for

additional details. An ISL mimics essential features of the in situ perfused rat liver systems used in

(Hung et al., 2001) to study the hepatic disposition of the cationic drugs. We assumed that rat, hepatic,

secondary unit function is similar throughout a normal liver, and that, on average, the function of different

lobules would be indistinguishable.

We therefore represented the entire liver as a collection of lobules. Blood flow is represented as

entering lobules via portal vein tracts and vascular septa, and draining into branches of a common central

vein (CV). The ISL represents each lobule as a network of cell-lined sinusoids connected by solute flows.

The network is a directed graph; each node (or vertex) represents a cell-lined sinusoid and a link (or edge)

between two nodes represents directional flow. For each LOBULE simulation, the graph structure is

determined stochastically within parameter-specified constraints. Objects representing sinusoidal spaces

and functions, along with related spatial features, are called "Sinusoidal Segments" (SSs). One SS is

placed at each graph node. The parameters that control LOBULE form and function for the experiments in

this report are listed in (Yan et al., 2007) along with all other ISL features. For convenience, they are also

listed in Table S1 of the Data Supplement. Only ten parameters were modified during the course of this

study. Except as noted below, all other parameter values were the same as those specified in Table S1.

The meaning of each parameter is in its name. For example, parameter *C2BJumpProb* is the probability

that a COMPOUND, when given the option, will move from Grid C to Grid B; parameter

BindersPerCellMax specifies the maximum number of binders that can be contained in any one CELL.

A SS is a tube-like structure with a blood "Core" surrounded by three identical size spaces in the form

of 2D grids. Grid A represents the sinusoid rim. Grid B, which represents endothelial cells and fenestra,

is wrapped around Grid A. Grid C, which represents all other spaces, including the space of Disse and

hepatocytes, is wrapped around Grid B. Different sinusoids can experience different flows and have

different surface-to-volume ratios. Such heterogeneity is represented by specifying two SS classes: S_1

and S_2. Compared to a S_2, a S_1 has a shorter internal path length and a smaller surface-to-volume ratio.

We represented the lobule interior as being subdivided into three concentric zones in order to

distinguish the quantitative difference in intralobular structural characteristics and, when needed,

enzymatic and transporter gradients. The number of directed graph nodes per zone is always in the order

of Zone 1 > Zone 2 > Zone 3. Each LOBULE zone contains at least one node. Because interconnections

between sinusoids are frequent in the periportal region and are rare near the CV, there are more

interconnections between Zone 1 nodes than between Zone 2 nodes. There are no interconnections

between Zone 3 nodes. All of these features are controlled by parameters (Table S1), many of which are

stochastic to simulate uncertainty and biological variability.

Simulation of Drug Disposition. ISL experiments follow the same protocol used for in situ perfused

rat liver experiments. Specifically, in (Hung et al., 2001), a bolus containing similar amounts of the

extracellular space marker sucrose and a cationic compound was injected into the portal vein through a

catheter (because of this, drug does not enter the liver as a bolus or a short duration square wave).

Consequently, the ISL dosage function is a modified gamma function (given in Table S1; see (Hunt et al.,

2006) for details) rather than an impulse to simulate effects of catheters and large vessels. Perfusate was

collected using a fraction collector. The fraction of the administered dose contained per collection

fraction was measured. Hepatic compounds of interest in the ISL are represented as COMPOUNDS (also

called DRUGS): they are objects that move through the LOBULE. Each COMPOUND represents a large number of drug molecules and can interact with each encountered SS feature. A COMPOUND's behavior is determined by the PCPs of its referent, along with the consequences of its interaction with the LOBULE and SS features encountered during its unique trek from portal vein tracts to CV.

After COMPOUNDS enter the LOBULE, they can enter a SS in Zone 1 at either the Core or Grid A. Thereafter, until each is collected at the CV, they have several stochastic options. The parameter *CoreFlowRate* biases COMPOUND movement in the Core in the CV direction, simulating blood flow. The parameter *SinusoidTurbo* biases COMPOUND movement in the three spaces in the CV direction, but much less so than *CoreFlowRate*; it simulates turbulent flow. In the Core or Grid A, a COMPOUND can move within that space, jump from one space to another, or exit the SS. When a COMPOUND jumps to Grid B, it can move within the space, jump back to the Grid A, or on to Grid C. All COMPOUNDS can move within the extracellular portions of Grids B and C. When a COMPOUND encounters a CELL in Grid B or C, it can move into it if allowed to do so by the CELL (it is allowed when the parameter *isMembraneCrossing* is true). When a CELL sees SUCROSE, it also sees that its value of *isMembraneCrossing* is false; consequently, it will not allow SUCROSE to enter. All CELLS in Grid C represent HEPATOCYTES. After a COMPOUND exits a SS in Zone 3, it enters the CV. Its arrival is recorded, simulating being collected by a fraction collector. An animated visualization of these events during a simulation, from start to finish, at both individual ISL and SS levels, for SUCROSE and ANTYPYRINE administered together, is available through Supplemental Data of (Yan et al., 2007).

The only subcellular functions that were needed to simulate behaviors of the six drugs under study

were binding and metabolism. As done in (Liu and Hunt, 2005; Liu and Hunt, 2006; Sheikh-Bahaei et al.,

2006), transporters can be added when needed. At the start of each simulation, objects functioning as

containers and representing CELLS are placed randomly at some fraction of the available locations in

Grids B and C. All cellular components that bind or metabolize drugs, such as transporter, enzymes and

other cellular materials, are conflated and represented by simple functional objects called BINDERS. Each

CELL contains a randomly specified (within parameter-controlled limits) number of BINDERS in a well-

stirred space. A BINDER within a Grid B CELL can only bind and later release a DRUG. A BINDER within

a Grid C HEPATOCYTE, referred to as an ENZYME, can bind DRUG and either release or METABOLIZE it. In

the later case, the DRUG vanishes; for this report, a METABOLITE replacement is not used.

Because there are 38 ISL parameters (Table S1) that specifically influence disposition events, a major

simulation task in (Yan et al., 2007) was to locate a region of ISL parameter space capable of providing

parameter vectors that generate biologically realistic ISL outflow profiles for SUCROSE and each of the

four DRUGS. To accomplish that, the LOBULE graph and SS structure was first tuned for SUCROSE; we

held those values constant and then tuned the ten PCP-sensitive parameters until we matched outflow

profiles for each of the four cationic drugs. For this study, the parameters in Table 1 were specified as

being sensitive to differences in PCPs. All 28 other ISL parameter values (PVs) are unchanged: they

control LOBULE and simulation features that were considered independent of administered COMPOUND

and are identical to those specified in (Yan et al., 2007). Five of the ten parameters are used only for

COMPOUNDS that can enter CELLS. We located a specific ISL parameter domain that could generate valid

outflow profiles for all five compounds. The values of the PCP-sensitive parameters provided a base

from which we could predict ISL PVs for two new simulated cationic drugs: PRAZOSIN and

PROPRANOLOL.

Due to the stochastic nature of ISL simulations, each in silico experiment generates a slightly

different outflow profile. In this report, the outflow profiles from 48 simulation trials are summed to

represent one experiment.

Mapping PCPs to ISL Parameter Values. Traditional parameter estimation methods (Ljung, 1998)

are often infeasible for synthetic, discrete, agent-oriented, biomimetic models of the ISL type, because

many of the mathematical assumptions underlying these traditional methods are not satisfied. Therefore,

in addition to simply stating and describing the methods used, we also state considerations and provide

the rationale.

It is widely held that a mapping exists between the hyperspace of drug PCPs and the hyperspace of

measured PK properties (Mager, 2006). Correlation results, as in (Hung et al., 2001), frequently

demonstrate relationships between a PCP and values of one or more fitted PK parameters, providing

evidence of such mappings. Validation of the ISL for five compounds provided evidence for mappings A,

B and C illustrated in Fig. 2: the mapping between A) ISL outflow profiles and their in situ counterparts;

B) ISL events and hepatic disposition events; and C) local ISL mechanisms and their drug-hepatic

component counterparts. Because a mapping exists between PCPs and the causal hepatic mechanisms,

we infer that a logical mapping can also be constructed between drug PCPs and the PCP-sensitive subset

of tuned, ISL parameters of the four compounds in Table 1.

How can we establish specific, quantitative mappings so that ISL PVs for a new drug can be approximated directly based on its PCPs? How can we accomplish the task when we have "only limited drug-specific data"? An example of how this can be done for a synthetic, discrete event, agent-oriented model having design features similar to the ISL, was recently presented (Sheikh-Bahaei and Hunt, 2006), and theirs is one of the methods that we use.

Methods for Predicting ISL Parameter Values. What is the best method and approach for predicting PCP-sensitive ISL PVs for a researcher-selected set of PCPs? We know from exploratory ISL experiments that the change in an outflow profile caused by changing the value of one parameter can be compensated by modest adjustments in the values of several other parameters: the relationships among ISL PVs and outflow profiles are nonlinear (Hunt et al., 2006). From accumulated PK knowledge, we also know that mappings between PCPs and traditional PK model PVs can also be nonlinear. When attempting to identify and unravel complicated mappings, more data is always preferred. Unfortunately, the reality in the PK domain is that because of the high costs of wet-lab experiments and given the combinatorial size of all the factors involved in any PK mechanism, there will always be a shortage of specific data for any particular sub-system. In addition, most PCPs have their own uncertainty issues. Such considerations make predicting ISL parameters challenging. Before suggesting an answer to the lead question, we also need to take into account the following two issues.

1) A goal for this class of models is to incorporate the parameter prediction method(s) into the simulation system so that the ISL and its components can automatically adjust their DRUG-component interaction rules, "on the fly," using the PCP information carried by each COMPOUND. This will allow the

ISL components to adjust their interactions with DRUGS even though the particular COMPOUND has not

been previously encountered, and so no validated interaction rule is available.

2) The ISL is not a static model. It is expected to evolve iteratively with continued use. We can

anticipate that components and interaction events will be invalidated and changed. For example, a

specific mechanistic detail may be invalidated when the set of targeted attributes is expanded by inclusion

of a new attribute (based on wet-lab observations from other experiments). In order to validate against

the expanded set of targeted attributes, a current mechanistic ISL feature may need to be replaced by a

revised, possibly more detailed version. Consequently, parameter prediction methods need to be easily

transferable to new components as the ISL evolves.

Research in quantitative (drug) structure-PK parameter relationships has considered many mapping

methods ranging from simple linear regression to partial least squares regression (for examples see (Hou

et al., 2006) and (Ng et al., 2004)) and nonlinear regression methods, including artificial neural networks

(e.g., see (Nestorov et al., 1998)). Each has advantages and limitations. For results of the more

sophisticated methods to exhibit statistical significance, larger data sets are required, often acquired

considering the nature and capabilities of the prediction method. We can use any of these methods as

long as we are cognizant of the limitations, especially when working with sparse data: over-fitting is an

important limitation, but it does not preclude cautious use of a method. Another is the sensitivity of the

model's behavior (e.g., the outflow profile) to changes in predicted PVs. The impact of such limitations

depends on the intended use to which the predicted outflow profile will be put. Precise, verifiable

predictions are beyond the scope of ISL class models (see discussions in (Hunt et al., 2006) and (Yan et

al., 2007)). At best, an ISL can be used to anticipate the behavior of a new compound in the context of

considerable uncertainty, given its assumptions. An example of the latter: the hypothesized mappings B

and C in Fig. 2 are acceptably realistic for the intended use and can be extended to include the new

COMPOUND. It becomes clear that ISL simulations made using predicted PCPs is a rough approximation

of what might be expected if all assumptions are valid. Further, this is not a stand-alone prediction: it

should be considered only relative to validated ISL behaviors.

With the full weight of the above caveats in mind, a logical strategy would be to use several

parameter prediction methods. Simulations made using each of the predicted parameter sets for the new

drug can then be compared to in situ outflow profile of that same drug. We can state the following.

Consider the hepatic disposition properties of prazosin and propranolol determined from wet-lab

experiments and analogous properties of the two in an in silico experiment, where the in silico and wet-

lab experimental designs are effectively the same. The relative differences and similarities in simulated

disposition properties between the two compounds would approximate those of the wet-lab experiment.

Our expectation is that the ISL will be revised and reused to represent the hepatic disposition

properties of an increasing number of drugs under a variety of experimental conditions. With

improvement, the two hypothesized mappings, B and C in Fig. 2, will become more realistic because

validated, finer granularity versions of the simulated mechanisms will be available. As the simulated

mechanisms become less abstract, we can expect a shrinking subset of PCPs (or molecular descriptors

calculated from structure information) will become increasingly important in determining the outcome of the

referent event. In that case, the differences between predicted, PCP-sensitive PVs, arrived at using

14

different methods may decrease. If so, the scientific usefulness of the simulations can be expected to

increase.

Many molecular descriptors and PCPs (measured and calculated) are available (for examples, see (Ng

et al., 2004)). However, motivated by the desire to keep new methods as simple as possible, while still

achieving the objective, we elected to focus initially on just the nine in Table 2, the same set used in (Yan

et al., 2007). We elected not to include pK_a values for this study set for the sake of simplicity: we wanted

to postpone dealing with the potentially complicated issues of multiple pK_a values and the variety of drug-

specific pK_a values that one can encounter in the literature.

Below, we describe how we used three fundamentally different parameter prediction methods.

Method 1: Correlation and Linear Regression. The simplest method for specifying a quantitative

mapping is to use correlation and linear regression. We correlated each of four sets of PCP values in

Table 2 against each of four sets of PCP-sensitive, ISL PVs in Table 1. Next, we selected the one

correlated pair having the largest correlation coefficient (CC). We then used the corresponding, least-

squares regression line to predict a value of that ISL parameter for both PRAZOSIN and PROPRANOLOL,

given the value of the selected PCP.

Method 2: Fuzzy c-Mean Clustering. To handle uncertainty that arises during mappings of this

type, we used the Fuzzy c-Mean algorithm (FCMA) introduced by Bezdek et al. (Bezdek et al., 1984; Pal

and Bezdek, 1995). It can be used to transform descriptive mappings into quantitative relationships. It

15

starts with influence relationships between subsets of PCPs and subsets of PCP-sensitive, ISL PVs from

validated ISLs.

Sheihk-Bahaei and Hunt (Sheikh-Bahaei and Hunt, 2006) used the FCMA and "only limited drug-

specific data" to predict parameter values for their analogue of sandwich-cultured hepatocytes. They

made the estimation method available to the model to enable it to reasonably anticipate (predict) the

biliary transport and excretion properties of a new compound, enkephalin, based on the acceptable,

already tuned PVs of salicylate, taurocholate, and methotrexate. By assuming the similarity among PCPs

extends to model PVs, acceptable values of the latter for enkephalin (with PCPs that are very different

from the three other drugs) were predicted that had good similarity measure values. They established a

single mapping between thirteen selected PCPs and fourteen PCP-sensitive model PVs (the ISL has just

ten). However, correlations between specific PCPs and specific PK parameter values (Mager, 2006)

suggest that a small set of PCPs will have a dominant influence on specific drug-component interactions

during disposition, an observation confirmed recently by Ng et al. (Ng et al., 2004). It thus seemed

reasonable to expect that different subsets of PCPs will map to the hepatic counterpart of different ISL

events, such as moving between Grids B and C and movement into CELLS. Further, the artificial neural

network method discussed below provides a mapping between all nine PCPs and all ten ISL PCP-

sensitive, PVs. Therefore, for this second prediction method, we elected to seek usable relationships

between subsets of PCPs and the ISL parameters controlling specific ISL events.

Consider the six compounds in Table 2. Each drug's nine PCP values, combined into a vector, can be

viewed as a single point in a nine-dimensional hyperspace. The six vectors or points (also referred to by

the name of the compound that the PCPs describe) can be classified into one-to-six groups based in part

on their closeness in PCP hyperspace. Any compound can be a member of more than one group: the

groups or clusters overlap. The researcher can specify the number of groups, or use the FCMA to search

for the optimal number. The FCMA assumes that each point has some degree of membership in each

group. It then assigns a membership degree value to every vector in each cluster. The FCMA inputs are:

1) the data set of n cases to be clustered (for our use, each case is a single drug, a vector of nine PCP

values); 2) a parameter m known as the fuzzy exponent; and 3) the number of user specified clusters c.

The first two of these inputs are discussed further in the following sections. As recommended in (Pal and

Bezdek, 1995), we always set $m = 2$. The output of a FCMA, illustrated in Fig. 3, is a c-by-n matrix,

containing the membership values of the fuzzy clusters.

ISL Parameter Categories. There are three categories of ISL parameters. Those that control lobule

graph and sinusoid structures are the same for all drugs. Those that control the PCP-sensitive parameters

comprise the third category. We need only predict the values of PCP-sensitive parameters. We divided

the ten PCP-sensitive ISL parameters in Table 1 into four subgroups: those that reflect the influence of a

COMPOUND'S A) effective size, B) movement between spaces, C) binding to LOBULAR components, and

D) METABOLISM.

We took the nine PCPs in Table 2 and divided them into five groups, as follows, according to their

perceived similarity in influence: I) MW and volume; II) partition coefficient ($\log P_{app}$); III) unbound

fraction (f_{uB}); IV) topological parameters: number of rotatable bonds (RB count) and topological polar

surface area (TPSA). These two PCPs are known to be good descriptors for characterizing drug

17

absorption, including intestinal absorption, bioavailability, Caco-2 permeability, and blood-brain barrier

penetration (Ekins et al., 2000; van de Waterbeemd and Gifford, 2003). And, V) hydrogen bond related

parameters: hydrogen bond donor (HBD) count, hydrogen bond acceptor (HBA) count, and tautomer

count. These three PCPs are known to be important to a compound's chemical interactions with other

solutes and migration capabilities in aqueous materials(Ekins et al., 2000; Hansch et al., 2004). Because

the relationship between PCPs and ISL parameters are believed to be non-linear, we conservatively

assumed that more than one set of PCPs could influence more than one set of ISL parameters. We then

made the following mapping assumptions (Fig. 4) and took them as specifications for moving forward:

1) $\log P_{app}$ is the primary PCP reflecting a drug's partitioning ability and so is expected to map to a

    compound's ability to move between spaces. We also assumed that parameters in PCP subgroup IV,

    which reflect drug flexibility and topology, also contribute to a compound's ability to move between

    spaces: in the ISL, that ability is controlled by ISL parameter subgroup B. We specify that a

    quantitative mapping can be established between PCP subgroups II and IV, and ISL subgroup B. The

    mapping is illustrated in Fig. 4.

2) In addition to f_{uB}, parameters in PCP subgroup V are expected to contribute to drug binding to (or

    sequestration in) various cellular components. We therefore specify that a quantitative mapping can

    be established between PCP subgroups III and V, and ISL subgroup C (see Fig. 4).

3) DRUG METABOLISM (ISL parameter group D) simulates a complicated process and so it may map to a

    broader range PCPs. We included the following: $\log P_{app}$ (subgroup II) along with the PCPs in

    subgroups IV and V.

4) Both MW and volume are known to influence a compound's effective size in biological systems

(Hansch et al., 2004). We specify that a quantitative mapping can be established between them, and

ISL2WetLabScaling.

Fuzzy Cluster Prediction Strategy. In this section, we explain how we used FCMA results to

construct quantitative mappings between the PCP and ISL parameter spaces, and how we used those

mappings to predict the PCP-sensitive, ISL PVs for PRAZOSIN and PROPRANOLOL. For clarity, we first

describe the prediction method using an example.

The properties in PCP subgroups II and IV are posited to map to the PCP-sensitive subgroup B

parameters (Fig. 4). We illustrate a method to calculate a value of *B2CJumpProb* for PRAZOSIN. The

subgroup II and IV values are listed in Table 3. We apply the FCMA to these five, three-value, parameter

vectors, specifying three fuzzy groups: we get the degree of membership in each cluster listed in Table 3.

We see that prazosin is primarily a member of a group that contains at least one of the four already-

studied drugs. In this case, it is within a group to which labetalol and diltiazem also primary members.

Atenolol and antipyrine are primarily in different groups. We assume that the position of prazosin in ISL

parameter space relative to the other four drugs is approximately the same as its position in PCP space

relative to the other four drugs. As a first approximation, we specify that it is the case. Given that

assumption, we can calculate a value of *B2CJumpProb* using PRAZOSIN'S PCP degree membership values

in each of the three clusters along with the average values of *B2CJumpProb* for all the members in each

of the three clusters. LABETALOL and DILTIAZEM are members of cluster 1. The average *B2CJumpProb*

value is 0.5; the value for the only primary member of cluster 2 (antipyrine) is 0.35; the value for the only

primary member of cluster 3 is 0.3. Given that data, for PRAZOSIN, we have

$$B2CJumpProb = (0.506)(0.5) + (0.179)(0.35) + (0.316)(0.3) = 0.41$$

Additional, detailed calculations are provided in the Data Supplement.

The same approach can be used to calculate the remaining three members of ISL subgroup B. The

method can also be extended to calculate expected values for the other ISL parameters. The generalized

statement of the method follows.

Parameter Estimation via Analysis of Fuzzy Clusters. 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*. For additional detail, see (Sheikh-Bahaei and Hunt, 2007). In

this work, the new situation is the ISL encountering two new types of objects, one representing prazosin

and the other representing propranolol. The algorithm starts with the highest possible number of clusters

and iteratively decreases the number of clusters until the new compound is placed in a cluster with at least

one of the already-tuned compounds. The compounds in that cluster represent those "most similar" to the

new compound.

In general, for a set *S,* containing *n data points* $S = \{c_1, c_2, \ldots c_n\}$ (one *data point* corresponds to one

situation—one drug, in our case, with its associated PCP values)*,* the following algorithm estimates the

simulation parameters of a new *situation*, c_{n+1}:

Step 1. Let $q = n$, and $S_{new} = \{c_1, c_2 \ldots c_n, c_{n+1}\}$.

Step 2. If $q = 1$ go to step 4. Else, classify S_{new} into q clusters using the FCMA.

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 group-mates of c_{n+1}

Step 4. Call the q groups G_1, G_2, … G_q, where $c_{n+1} \in G_1$. Let μ_k be the membership degree of c_{n+1} in G_k.

Estimate the needed ISL simulation parameters of c_{n+1} as:

$$\hat{\vec{P}}_X = \sum_{k=1}^{q} \mu_k \cdot \vec{P}_{Gk} \qquad (1)$$

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

$$\hat{\vec{P}}_{Gk} = (\sum_{j=1}^{m} \mu_j \cdot \vec{P}_j) / \sum \mu_j \qquad (2)$$

The accuracy or usefulness of the resulting estimates depends, of course, on how many *data points*

similar to c_{n+1} exist in the data set, i.e. the higher the *G-value*, the better the accuracy. Because we are

starting with just four compounds, four data points, our expectations must necessarily be modest.

Method 3: Artificial Neural Network (ANN). To establish a traditional nonlinear mapping, we

used a simple feed-forward ANN having a single hidden node (Hudson and Cohen, 1999). The neuron

transfer function is $y(x) = b/(1 + e^{-ax})$; we chose $a = b = 1$. We recognized that the small training set

would lead to an over-fitted prediction model. Nevertheless, we used it because the method is known to

be good at finding patterns and establishing quantitative relationships between data sets based on those

patterns. We also used it to provide an indication of what might be expected from an ANN-based,

predictive mapping, recognizing that such mappings have often done better than one of the many

available multivariate linear mapping options, when the relationships between the data sets are known to

be complicated, as is the case here. We used a backpropagation training algorithm that used normalized

values of the four sets of nine PCPs in Table 2 as inputs; values of the four sets of ten ISL PVs in Table 1

were obtained as outputs. We then used the trained ANN to separately predict a full set of ten PCP-

sensitive, ISL PVs for PRAZOSIN and PROPRANOLOL, given their full set of nine PCPs as inputs.

Values were normalized as follows: divide each point by its vector Norm. E.g., if drug (point) X has

PCPs *A, B* and *C*, then it was divided by its length *N*, i.e., $N^2 = A^2 + B^2 + C^2$ and $X_{normalized} = (A/N, B/N, C/N)$. The output values were normalized in the same way. Consequently, the predictions too were

normalized; to be used, the predictions had to be scaled: multiplied by the norm of the training vector.

Similarity Measure. The Similarity Measure (SM) used to compare in silico with the lag-time

adjusted, referent outflow profiles is detailed and discussed in (Hunt et al., 2006) and (Yan et al., 2007).

Briefly, we assumed that the coefficients of variation of identically timed observations between t = 1 and t

= 100 seconds from repeat experiments (referent) are constant. For each referent outflow profile measure

P, we create two curves, $P^l = P(1 - d)$ and $P^u = P(1 + d)$. They form the lower and upper bounds of a band

around *P*. The SM value is the fraction of evenly spaced, outflow profile measures falling within the

band. Here, *d* is the standard deviation (SD) of the relative differences between each of six repeated in

situ experiments and the mean value for a given collection interval, pooled over all collection intervals.

In (Yan et al., 2007), where the ISL outflow profile was iteratively tuned to give a good match to the

referent profile, we used *d* = 1 (SD), and an ISL outflow profile was deemed similar to the referent if 80%

or more of ISL outflow values were within the band (SM ≥ 0.8). Here, using predicted PVs without any

post-prediction refinement, we speculated that having a predicted profile fall within a factor of two of the

referent profile could be considered a good result. Nevertheless, we calculated SM values using two

reference bands: $d = 1$ and 1.5, which corresponds to factors of 0.45 and 0.675.

ISL Execution. Experiments were executed on an eight node OSCAR cluster

(oscar.openclustergroup.org/) running RedHat's Fedora 5. The distribution of the runs uses MPICH 1.2.7

(www-unix.mcs.anl.gov/mpi/mpich1/). The ISL was compiled using GCC 4.1.1 against the Swarm 2.2.3

Objective-C libraries (swarm.org/wiki/Main_Page). The initial PRN seed was extracted from the

machine's clock. Each completed experiment was archived and logged with the date and time using a

Makefile target. Examination and processing of data from simulations used a combination of Matlab

(7.0.0) and R (2.4.0) (www.r-project.org/) scripts, and Microsoft Excel. As done previously (Hunt et al.,

2006) (Yan et al., 2007), we smoothed results from each simulation experiment using the

Smodulus_smoothing function from the Rwave (version 1.22) package of R. For the results presented, a

wavelet window of three observations at 0.5-second intervals was used.

Results

Predicted PCP-sensitive ISL Parameter Values. Ninety correlations were performed: four values

of each of ten ISL PVs (from having already tuned the ISL for four drugs) with the four values of each of

nine PCPs corresponding to those four drugs. Table 4 lists the CC values. The regression line for the ten

pairs having the largest CCs (bold) was used to predict the ISL parameter value for both PRAZOSIN and

PROPRANOLOL listed in Table 5.

The PCP-sensitive ISL PVs for PRAZOSIN and PROPRANOLOL predicted using Eqs. 1 and 2 in combination with the mappings specified in Fig. 4 are also listed in Table 5, along with the ISL PVs for the two drugs predicted using the ANN. Because the referent outflow profiles for prazosin and propranolol included co-administered sucrose, equations 1 and 2 were also used to calculate new PCP-sensitive ISL PVs for SUCROSE. Those values are also listed in Table 5. Several examples of using the FCMA method to predict specific ISL parameter values are provided in the Data Supplement along with the predicted sucrose outflow profiles (Fig. S1) when SUCROSE was co-administered with PRAZOSIN and PROPRANOLOL.

Hepatic Disposition Predictions. We used the predicted ISL PVs in Table 5 along with the previously specified, PCP-independent, ISL parameter values (Table S1) to parameterize new ISLs, three for PRAZOSIN and three for PROPRANOLOL. Simulations using each of these ISLs generated the expected outflow profiles presented in Figs. 5-7. One simulated outflow profile is the result of pooling 48 separate Monte-Carlo trials. One trial represents a single lobule and 48 trials represent the whole liver.

As previously discussed (Yan et al., 2007), the structural and microarchitectural details of the ISL used for each simulation trial were nondeterministic. Even though the number of nodes per zone and the number of intra- and interzone edges were specified, the connectivity pattern was determined randomly at the start of each simulation. SS structures for a given set of parameter values are also stochastic: the actual structures of the 72 SSs used for each ISL run were highly constrained yet different. Because of theses differences, combined with the many forms of discretization, there was variability between ISL instantiations and, consequently, the raw outflow profiles were noisy. That variability reflects, and to

24

some extent simulates, interindividual hepatic differences along with the uncertainty in our knowledge

about the exact details of the generative mechanisms. Increasing the number of averaged simulation runs

makes outflow profiles smoother. The additional wavelet smoothing was described under Methods.

As described under Similarity Measure, when matching the simulated outflow profile with lag-time

adjusted, *in situ* outflow profiles, we used a band of ±1 SD flanking the original wet-lab outflow profile

values as a measure of the similarity of ISL simulations to those wet-lab results. For the outflow profiles

generated from the predicted parameter values, two reference bands were used: ±1 SD and ±1.5 SD, and

SM values were calculated for each. For these data, a band of ± 1 SD corresponds to a constant, 45%

coefficient of variation; ± 1.5 SD corresponds to a constant, 67.5% coefficient of variation. The ±1 SD

SM values, in the order PRAZOSIN and PROPRANOLOL, shown in Table 5, were as follows. Method 1

(simple linear regression): 0.36 and 0.68; Method 2 (the FCMA): 0.50 and 0.58; and Method 3 (ANN):

0.79 and 0.97. The corresponding ±1.5 SD SM values: Method 1: 0.55 and 0.82; Method 2: 0.77 and 0.72;

and Method 3: 0.97 and 0.98.

Drug-specific Prediction Strategies. Although the predicted ISL parameter values from the FCMA

method generated good predicted PRAZOSIN and PROPRANOLOL outflow profiles, they were the result of

applying simultaneously a common PCP-to-ISL mapping strategy to both drugs. We recognize that

prazosin and propranolol have significantly different PCPs. What if we had sought parameter predictions

for each drug separately? Would individualized mappings yield improved outflow profile predictions?

The mapping between each compound's PCPs and the matched ISL parameter values are not expected to

be identical. We explored the consequences of individualizing the mappings. Several dozen mappings

were explored. Starting with the strategy in Fig. 4, we made one mapping change, then predicted

parameter values, and simulated outflow profiles. We then repeated the process for a second, third, etc.

mapping change. For both drugs, we kept the number of PCP-ISL mappings constant at four, for four

subgroups of ISL parameters, as in Fig. 4. We removed or added single PCP values or single PCP groups.

Most of the mapping options resulted in expected drug outflow profiles having SM values for both

PRAZOSIN and PROPRANOLOL that were smaller than those in Table 5 for the FCMA method. In several

cases, the SM values showed little change. Only a few showed improved SM values, but they were

modest. The individualized mappings for two of the latter cases (Fig. S2), along with the corresponding

simulated outflow profiles (Fig. S3) are provided in the Data Supplement.

Discussion

Prediction of PK parameter values plays an increasingly important role in decision making

concerning drug candidates in development. The methods and issues, which have been thoroughly

reviewed, fall into two categories: empirical, data-based approaches (Stouch et al., 2003; Beresford et al.,

2004; Yamashita and Hashida, 2004; Li et al., 2007) and structure-based approaches (Ng et al., 2004; Li et

al., 2007). Data-based approaches can be subdivided into linear and non-linear methods, and each

approach may involve clustering. Examples of the former include multiple linear regression, partial least

squares analysis, and linear discriminant analysis. Examples of the latter include ANNs, genetic

algorithms, support vector machines, and k-nearest neighbor algorithms. Popular clustering techniques

include partitional algorithms (including k-Means Clustering), nearest neighbor clustering, ANNs,

evolutionary approaches, search-based approaches (including simulated annealing), and fuzzy clustering.

We employed clustering, specifically fuzzy clustering, as one of the three methods used because it is

difficult to classify PCPs and the drug's they characterize into crisp, disjoint clusters. Conventional

clustering produce partitions: each pattern fits in one and only one cluster. Fuzzy clustering assigns each

pattern within each cluster a membership value. We speculated that this feature may make it an effective

tool for clustering PCPs for mapping to ISL PVs, especially when data are sparse.

The relative merits of linear and non-linear methods when dealing with traditional PK parameter

estimation are now reasonably well understood. However, the ISL is representative of a new class of

models. It contains more mechanistic detail than traditional PK models, and only a sub-set of its

parameters is PCP-sensitive. Consequently, we had no a priori basis for selecting one of the preceding

methods over another. We chose three quite different methods to begin building insight into their relative

merits for parameter estimation within the ISL context, given the multiple levels of uncertainty.

We started with a simple version of each method. We treated the predictions arising from their use as

semi-independent "votes" for how to parameterize an ISL, given a set of PCPs of prazosin and

propranolol, along with ISL PVs that are independent of PCPs from having validated ISLs for sucrose,

altenolol, antipyrine, labetalol, and diltiazem. Simulations arising from use of the three different sets of

predicted ISL parameter values gave us different hepatic disposition and outflow predictions. Because

referent data were sparse, we were not trying to select a best prediction method. We took each outflow

prediction in Figs. 5–7 as a plausible estimate what one might expect following administration of prazosin

and propranolol to a perfused liver. In the case of the ANN (Fig. 7), we recognized the risks associated

with over-parameterization (Hudson and Cohen, 1999). All three methods gave rise to predicted profiles that were within a factor of two of the actual profiles. We found it noteworthy that the simple, single PCP correlation method (Fig. 5) lead to ISL outflow profiles that were close to those of FCMA and ANN.

A single point in discretized PCP space represents an unique compound. The space is not continuous. Each point can map to a relatively unique hepatic outflow profile, which is a consequence of many entangled local, intralobular mechanisms. A mapping from PCP space to a disposition outflow profile (an attribute of the drug-liver phenotype) is complex. A small shift in location in PCP space is expected to correspond to distinguishable outflow profile changes. Table 2 shows that for this work, PCP space has nine dimensions. It is relatively simple to calculate dozens of additional physicochemical descriptors, ranging from simple geometric to sophisticated electrostatic and thermodynamic. On the other hand, with larger data sets, one can reduce the number of dimensions using a linear dimensionality reduction method such as Principal Component Analysis. However, the goal of this work is not finding the optimal size of PCP space. Rather, it is to explore the usefulness of the new methods in this context.

The ISL is an assembly of componentized mechanisms: purposefully separated and abstracted aspects of hepatic form, space, and organization interacting with COMPOUNDS. Each component mechanism has been unraveled from the complex whole of the hepatic-drug phenotype; it has its own unique phenotype, but that phenotype is much simpler than that of the entire lobule mechanism. Hence, a mapping from component behavior space to PCP space is also simpler and fundamentally different from the PCP-outflow profile mapping described above. For each ISL component, a small location change in PCP space may correspond to a negligible or modest change in the properties of that component mechanism,

as well as to the value of the parameter controlling its behavior. We see evidence of this in Tables 1 and

5: there are cases where different compounds have similar ISL PVs. Note also that for both PRAZOSIN

and PROPRANOLOL, there are cases where each of the three methods predicted similar PVs for the same

drug. Therefore, considered in retrospect, it is not surprising that even an overly simple estimation

method can lead to an acceptable parameterization. Interestingly, the current level of mechanistic

granularity is sufficiently fine-grained so that the subtle patterns in PCPs and in the PCP-sensitive, ISL

PVs that are needed to achieve larger SM values (Fig. 7; Table 5) are just enough for the simple ANN

parameter estimation technique, the more complicated of the three, to find them.

The results suggest that when using the synthetic method of assembling separated aspects of

mechanism, a parameter estimation method, which reasonably quantifies the relative differences between

compound-specific behaviors at the level of detail represented by those mechanisms, will provide

satisficing (the minimum satisfactory outcome), ballpark estimates. That bodes well for using the ISL

class of models for predicting PK properties, given only molecular structure information. However, the

six drugs in this study are all weak bases. As such, they occupy a common sub-region of PCP space.

More drug cases, spanning additional regions, will be needed to build confidence that the above,

apparently favorable parameter prediction situation is a consequence of the synthetic method and its

instantiation for the ISL (rather than fortuity).

Because it is a synthetic model, the ISL successfully shrinks the PCP-sensitive phenomenal manifold

relative to that of tightly coupled mathematical PK models. Traditional equation-based, PK models

reduce the phenomenal manifold to a set of real-valued parameters interacting within and assuming a

relational continuum. Although one can reduce the number of parameters being considered with these

models, in order to maintain accurately quantified relationships, the set of reduced parameters must still

accurately capture the aggregated relationships of the larger parameter set. This implies that, for tightly

coupled PK models, the phenomenal manifold remains just as complex even though there are fewer

parameters in the PK equations. The parsimony achieved by such reduction is, in a sense, illusory. The

fine-grained relationships that traditional PK models conflate into mathematical parameters are deeply

embedded in the coarse-grained relationships of these parameters. In contrast, the ISL can be actually

simplified because it is an assemblage of coupled, abstract mechanisms, such as transfer between spaces

and entry into cells, where the fine-grained relationships can be completely ignored.

Of course, a negative consequence of the ISL's componentized simplification is that the model can

sometimes yield dramatically inaccurate and abiotic simulations. Here are three examples: 1) a

parameterization makes too much space available to DRUGS causing a prolonged, abiotic lag-time to occur

before a slowly rising wave of DRUG emerges. 2) When the diversity in travel paths becomes too limited,

travel paths can form distinct clusters, rather than being smoothly distributed. That discontinuity in path

options can cause abiotic, oscillations to occur in outflow profiles. Evidence of oscillations can be seen in

the PROPRANOLOL profile prediction in Fig. 6. 3) Inaccurate and abiotic crowding can occur when too

many DRUGS converge too quickly on a too small space; this can occur when too many edges connect to

one SS. However, *all models are inaccurate*. A central purpose to the synthetic method is to clearly

identify and denote both the accurate and inaccurate assumptions and behaviors that characterize a model.

Similar phenomena occur with all wet-lab models. Aspects of *in vitro* phenotype are expected to map to

corresponding attributes in animal species or patients. However, combinations of experimental

specifications can cause *in vitro* behaviors and phenomena for which there is no counterpart in the

referent.

As long as we maintain awareness of the above negative consequence, such effects can be minimized.

For research that adopts the motivating vision of PBPK, it seems important to use both the synthetic

method for chronicling modeling assumptions (heuristic value) and the inductive method to achieve high

tolerance predictive value. We posit that, by having simplified, yet fine-grained, heuristic models, like

the ISL, which are commensurate with relatively simple parameter estimation methods, the biological

mechanisms become more understandable and accessible. Increased accessibility can provide the

methodological leverage needed to build synthetic analogues that will enable us to explore, digest,

categorize, understand, and use the massive amounts of data being generated by "omic" technologies.

Acknowledgements

We thank our collaborator Michael S. Roberts for providing us with the original liver perfusion data; we

also thank Prof. Roberts along with the other members of the BioSystems Group for helpful discussion

and commentary.

References

Beresford AP, Segall M and Tarbit MH (2004) In silico prediction of ADME properties: are we making progress? *Curr Opin Drug Discov Devel* **7:**36-42.

Bezdek JC, Ehrlich R and Full W (1984) FCM: The fuzzy c-means clustering algorithm. *Computers and Geoscience* **10:**191-203.

Ekins S, Waller CL, Swaan PW, Cruciani G, Wrighton SA and Wikel JH (2000) Progress in predicting human ADME parameters in silico. *J Pharmacol Toxicol Methods* **44:**251-272.

Ertl P, Rohde B and Selzer P (2000) Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties. *J Med Chem* **43:**3714-3717.

Hansch C, Leo A, Mekapati SB and Kurup A (2004) QSAR and ADME. *Bioorg Med Chem* **12:**3391-3400.

Hou T, Wang J, Zhang W, Wang W and Xu X (2006) Recent advances in computational prediction of drug absorption and permeability in drug discovery. *Curr Med Chem* **13:**2653-2667.

Hudson DL and Cohen ME (1999) *Neural Networks and Artificial Intelligence for Biomedical Engineering (IEEE Press Series on Biomedical Engineering)*. Wiley-IEEE Press.

Hung DY, Chang P, Weiss M and Roberts MS (2001) Structure-hepatic disposition relationships for cationic drugs in isolated perfused rat livers: transmembrane exchange and cytoplasmic binding process. *J Pharmacol Exp Ther* **297:**780-789.

Hunt CA, Ropella GE, Yan L, Hung DY and Roberts MS (2006) Physiologically based synthetic models of hepatic disposition. *J Pharmacokinet Pharmacodyn* **33:**737-772.

Li H, Yap CW, Ung CY, Xue Y, Li ZR, Han LY, Lin HH and Chen YZ (2007) Machine learning approaches for predicting compounds that interact with therapeutic and ADMET related proteins. *J Pharm Sci* **96:**2838-2860.

Liu Y and Hunt CA (2005) Studies of intestinal drug transport using an in silico epithelio-mimetic device. *Biosystems* **82:**154-167.

Liu Y and Hunt CA (2006) Mechanistic study of the cellular interplay of transport and metabolism using

the synthetic modeling method. *Pharm Res* **23:**493-505.

Ljung L (1998) *System Identification: Theory for the User (2nd Edition)*. Prentice Hall PTR.

Mager DE (2006) Quantitative structure-pharmacokinetic/pharmacodynamic relationships. *Adv Drug Deliv Rev* **58:**1326-1356.

Nestorov I, Aarons L and Rowland M (1998) Quantitative structure-pharmacokinetics relationships: II. A mechanistically based model to evaluate the relationship between tissue distribution parameters and compound lipophilicity. *J Pharmacokinet Biopharm* **26:**521-545.

Ng C, Xiao Y, Putnam W, Lum B and Tropsha A (2004) Quantitative structure-pharmacokinetic parameters relationships (QSPKR) analysis of antimicrobial agents in humans using simulated annealing k-nearest-neighbor and partial least-square analysis methods. *J Pharm Sci* **93:**2535-2544.

Pal N and Bezdek J (1995) On cluster validity for the fuzzy c-means model. *IEEE Trans Fuzzy Syst* **3:**370-379.

Rowland M, Balant L and Peck C (2004) Physiologically based pharmacokinetics in drug development and regulatory science: a workshop report (Georgetown University, Washington, DC, May 29-30, 2002). *AAPS PharmSci* **6:**E6.

Sheikh-Bahaei S and Hunt CA (2006) Prediction of in vitro hepatic biliary excretion using stochastic agent-based modeling and fuzzy clustering, in: L. F. Perrone, et al., eds., *Proceedings of the 37th conference* on *Winter simulation,* Monterey, CA, Dec 03 - 06, 1617-24.

Sheikh-Bahaei S and Hunt CA (2007) In silico hepatocyte: agent-based modeling of the biliary excretion of drugs in vitro, in: (L. Yilmaz, et al., eds.) *Proc 2006 Agent-Directed Simulation Symposium (Huntsville, AL, Apr 3-9)*, SCS Press, San Diego, CA (ADS'06), 157-63.

Sheikh-Bahaei S, Ropella GEP and Hunt CA (2006) Parameter Estimation via Analysis of Fuzzy Clusters (PEAF): An Algorithm to Estimate Parameters of Agent-Based Models, in: (G. Hu, ed.) *Proceedings of the 20th International Conerencef on Computer Applications in Industry and*

Engineering, ISCA, Cary, NC (ISBN: 978-1-880843-65-9), 23-7.

Stouch TR, Kenyon JR, Johnson SR, Chen XQ, Doweyko A and Li Y (2003) In silico ADME/Tox: why

models fail. *J Comput Aided Mol Des* **17:**83-92.

van de Waterbeemd H and Gifford E (2003) ADMET in silico modelling: towards prediction paradise?

Nat Rev Drug Discov **2:**192-204.

Yamashita F and Hashida M (2004) In silico approaches for predicting ADME properties of drugs. *Drug*

Metab Pharmacokinet **19:**327-338.

Yan L, Ropella GEP, Park S, Roberts MS and Hunt CA (2007) Modeling and Simulation of Hepatic Drug

Disposition Using a Physiologically Based, Multi-Agent In Silico Liver. *Pharm Res* Published

online: 28 November 2007 (DOI: 10.1007/s11095-007-9494-y)

DMD #19067

Footnotes

Figure Legends

Fig. 1. Illustrations of ISL model components. The six levels of ISL organization, from largest to

smallest, are: 1) the ISL experiment system, 2) SINUSOID network, 3) SINUSOID segment, 4) SINUSOID

spaces, 5) CELLS and EXTRACELLULAR domains, 6) ENZYMES and binders. Upper left: a schematic of a

cross-section of a hepatic lobule. Blood flow enters lobules via terminal portal vein tracts, goes through

sinusoids in three concentric zones, and drains into branches of a common central vein (CV). The three

different zones have quantitative differences in structural characteristics and DRUG interaction capabilities.

Upper right: objects representing drug and sucrose are injected as a bolus into a simulated catheter that

feeds into the portal vein tracts. Middle-right: a portion of the SINUSOID network is represented by an

interconnected, three zone, directed graph. Objects called SINUSOID segments (SSs) are placed at each

graph node: they simulate sinusoid functions. Lower-left: a schematic of SS component organization:

Grids represent spaces and can contain objects representing the functions associated with a portion of

SINUSOID space. Objects representing DRUG and SUCROSE enter and exit a SS via the Core and Grid A.

From Grid A, a COMPOUND can access the other spaces. Grid locations have properties that govern their

interaction with mobile COMPOUNDS. The potential for heterogeneous properties within different grids is

illustrated by different shadings of grid locations. Objects functioning as containers (for other objects) are

used to represent cells, and can be assigned to any grid location. The Core represents blood flow; Grid A

represents the sinusoid rim; Grid B represents endothelial cells and fenestra; Grid C represents all other

spaces, including the Space of Dissé and hepatocytes. Lower-right: A HEPATOCYTE container: objects

36

representing all needed intracellular components can be placed within. All enzymes that metabolize drug

are represented by ENZYME objects (e.g., CYP); cell components that simply bind, sequester, and release

drugs are represented by BINDERS. Additional binders representing transporters and other capabilities can

be easily added as needed. Cell containers can distinguish between DRUG and SUCROSE; they do not

allow SUCROSE to enter. Bile was not needed for these simulations, but it can be represented easily as an

extension, as described in (Sheikh-Bahaei and Hunt, 2006).

Fig. 2. Illustration of relationships between ISL mechanisms and its components, along with

relationships between the ISL and referent, perfused liver counterparts. To the extent that mappings A, B

and C are realistic, the relationship between a drug's physicochemical properties (PCPs) and specific

hepatic disposition events will have an ISL counterpart. Discovering plausible methods to make a

quantitative mapping between a drug's PCP and PCP-sensitive ISL parameter values will allow prediction

of DRUG–specific, PCP-sensitive ISL parameter values, given only a new drug's PCPs. ISL simulations

using those predicted parameter values stand as a prediction of the expected disposition behavior of that

new drug.

Fig. 3. Inputs and output of the Fuzzy c-Means algorithm (FCMA). The FCMA locates c clusters within

the set of n points (corresponding to n drugs) in PCP hyperspace and assigns a membership degree, μ_{ij},

for each point i to each cluster j.

Fig. 4. Mappings between PCP subgroups and ISL parameter subgroups. The four ISL subgroups are specified in Table 1; the five PCP subgroups are specified in Table 2. Each line represents a hypothesized strong influence. E.g., the partitioning and topology related PCPs are thought to have a strong influence on the wet-lab counterpart of the four PCP-sensitive ISL parameters values controlling movement between spaces within SSs. The identified PCP subgroups are the ones used by the FCMA to provide the membership degree values used in Eq. 1 to predict ISL PVs for PRAZOSIN and PROPRANOLOL in Table 5. The FCMA treats the identified PCPs as having equal influences. The influence of the other PCPs is neglected.

Fig. 5. Expected outflow profiles (black circles) for prazosin and propranolol using predicted parameter values. These semilog plots show results of simulation experiments for which the PCP-sensitive PVs were predicted using the Simple Linear Regression Method values listed in Table 5. All other ISL parameter values were identical to those reported in (Yan '07). Open squares: observed, referent wet-lab data. The data are the fraction of dose per outflow unit (per ml for the referent) as a function of time (1 simulation time unit = 1 second) after dosing with DRUG. The light gray band spans the range for the mean of the referent data $\pm$ one standard deviation (SD) as specified in the text. The dark gray band spans the range for the mean of the referent data $\pm$ 1.5 SD. Each ISL datum is the smoothed (window size: 3 values) mean value of 48 independent ISL runs.

Fig. 6. Expected outflow profiles (black circles) for prazosin and propranolol using predicted parameter

values. These semilog plots show results of simulation experiments for which the PCP-sensitive PVs

were predicted using the FCMA values listed in Table 5. Otherwise, the graph components are same as

specified in Fig. 5.

Fig. 7. Expected outflow profiles (black circles) for prazosin and propranolol using predicted parameter

values. These semilog plots show results of simulation experiments for which the PCP-sensitive PVs

were predicted using the ANN Method values listed in Table 5. Otherwise, the graph components are

same as specified in Fig. 5.

TABLE 1

Tuned PCP-sensitive ISL parameter values for four cationic drugs

ISL Attribute or Event	ISL Parameters	Atenolol	Antipyrine	Labetalol	Diltiazem
A. Effective DRUG size	*ISL2WetLabScaling*[a]	7	7	6	1
B. Movement between	*A2BJumpProb*	0.1	0.1	0.35	0.9
spaces	*B2AJumpProb*	0.6	0.6	0.2	0.2 [b]
	B2CJumpProb	0.3	0.35	0.5	0.5
	C2BJumpProb	0.6	0.65	0.5	0.2
C. Binding to LOBULAR	*BindersPerCellMin*	5	5	5	10
components	*BindersPerCellMax*	10	10	10	20
	SoluteBindingProb	0.35	0.5	0.6	0.35
	SoluteBindingCycles	25	25	25	20
D. METABOLISM	*MetabolizeProb*	0.35	0.4	0.3	0.02

[a] *ISL2WetLabScaling* = 7 for SUCROSE when administered alone or in combination with any drug.

[b] *B2AjumpProb* =0.35 for SUCROSE when co-administered with DILTIAZEM (compensates for difference

ISL2WetLabScaling values).

TABLE 2

Physicochemical properties of sucrose and six cationic drugs

Category	Parameters	Suc-rose	Aten-olol	Anti-pyrine	Prazo-sin	Labeta-lol	Pro-pranolol	Dilti-azem
I: Size	MW	342.3	266.3	188.2	383.4	328.4	259.4	414.5
	Volume	283.6	260.9	178.3	336.0	314.8	257.8	377.8
II: Partitioning	$\log P_{app}$ [a]	–	0.14	0.33	1.88	2.69	3.10	3.53
III: Binding	f_{uB} [b]	–	0.74	0.60	0.54	0.52	0.45	0.28
IV: Topology related	RB count [c]	5	8	1	4	8	6	7
	TPSA [d]	189.5	84.6	26.9	107.0	95.6	41.5	59.1
V: Hydrogen bound related	HBD count [e]	8	3	0	1	4	2	0
	HBA count [f]	11	4	2	8	4	3	5
	Tautomer count [g]	1	2	0	3	7	0	0

[a] $\log P_{app}$: log of the octanol-water partition coefficient at pH 7.4, 37ºC

[b] f_{uB}: Unbound fraction of each drugs in the liver perfusion medium as determined in (Hung et al., 2001)

[c] RB: number of rotatable bonds, a measure of molecular flexibility; amine C-N bonds are not considered because of their high rotational energy barrier

[d] TPSA: topological polar surface area is a sum of fragment contributions; O- and N- centered polar fragments are considered (Ertl et al., 2000)

[e] HBD: hydrogen bond donor

[f] HBA: hydrogen bond acceptor

[g] Tautomers are organic compounds that are interconvertible by tautomerization.

TABLE 3

Values used in the text example for predicting a value of B2CJumpProb for prazosin

Top: values comprising PCP subgroups II and IV. Middle: the results of applying the FCMA to the five

sets of three PCP values. Bottom: tuned *B2CJumpProb* values for the four already-studied drugs along

with the predicted value for prazosin.

PCP Subgroups II & IV	Atenolol	Antipyrine	Labetalol	Diltiazem	Prazosin
RB count	8	1	8	7	4
TPSA	84.58	26.94	95.58	59.09	106.96
$\log P_{app}$	0.14	0.33	2.69	3.53	1.88
Cluster					
C 1	0.007	0.001	**0.930**	**0.895**	**0.506**
C 2	0.006	**0.998**	0.020	0.045	0.179
C 3	**0.987**	0.001	0.050	0.060	0.316
ISL Parameters					
B2CJumpProb	0.3	0.35	0.5	0.5	(0.41)

DMD #19067

TABLE 4

Correlation coefficients between each PCP-sensitive, ISL parameter and each PCP for atenolol, antipyrine,

labetalol, and diltiazem

The least squares regression line corresponding to the largest correlation coefficient (bold) was used to predict

that ISL parameter value for PRAZOSIN and PROPRANOLOL, given their value of that PCP. Those ten, predicted

parameters values are listed in Table 5 and were used to generate the outflow profiles in Fig. 5.

	ISL2WetLabScaling	*A2BJumpProb*	*B2AJumpProb*	*B2CJumpProb*	*C2BJumpProb*
MW	0.87	0.92	0.87	0.79	**0.94**
Volume	0.82	0.88	0.87	0.77	0.91
$\log P_{app}$	0.83	0.91	**0.98**	**0.97**	0.89
TPSA	0.06	0.04	0.41	0.25	0.10
RB	0.28	0.34	0.50	0.34	0.42
f_{uB}	**0.93**	**0.95**	0.81	0.85	0.92
HBD	0.46	0.36	0.14	0.02	0.31
HBA	0.71	0.75	0.69	0.55	0.80
Tautomer	0.31	0.16	0.44	0.38	0.16

	BindersPerCellMin	*BindersPerCellMax*	*BindingProb*	*BindingCycle*	*MetabolizeProb*
MW	0.80	0.80	0.27	0.80	**0.92**
Volume	0.75	0.75	0.25	0.75	0.88
$\log P_{app}$	0.73	0.73	0.08	0.73	0.84
TPSA	0.16	0.16	0.12	0.16	0.06
RB	0.20	0.20	0.20	0.20	0.40

43

(Table 4, continued)

f_{uB}	**0.88**	**0.88**	0.08	**0.88**	0.90
HBD	0.57	0.57	0.40	0.57	0.36
HBA	0.66	0.66	0.43	0.66	0.80
Tautomer	0.45	0.45	**0.70**	0.45	0.23

TABLE 5

Predicted ISL parameter values for sucrose, prazosin, and propranolol

For each of the three methods, the listed parameter values were combined with the PCP-insensitive parameter values. The resulting ISLs were used to generate the expected outflow profiles in Figs. 5–7.

ISL parameter values	Method 1: Simple Linear Regression		Method 2: FCMA			Method 3: ANN	
	PRAZOSIN	PROPRANOLOL	SUCROSE[a]	PRAZOSIN	PROPRANOLOL	PRAZOSIN	PROPRANOLOL
ISL2WetLabScaling	5.3	4.1	6.0	3.0	7.0	4.6	2.8
A2BJumpProb	0.35	0.52	0.35	0.37	0.86	0.62	0.75
B2AJumpProb	0.37	0.21∗	0.22	0.40	0.21∗	0.29	0.21∗
B2CJumpProb	0.42§	0.50∗	0.49	0.41§	0.50∗	0.50§	0.50∗
C2BJumpProb	0.56§	0.31	0.50	0.48§	0.22	0.42§	0.32
MetabolizeProb	0.34	0.13∗	-	0.25	0.12∗	0.19	0.09∗
BindersPerCellMin	6	12	-	7	6	8	8
BindersPerCellMax	7	14	-	14	13	15	17
SoluteBindingProb	0.47§	0.39∗	-	0.41§	0.42∗	0.43§	0.42∗
SoluteBindingCycles	24§	23∗	-	23§	24∗	24§	23∗
SM values (±1 SD)	0.36	0.68		0.50	0.58	0.79	0.97
SM values (±1.5 SD)	0.55	0.82		0.77	0.72	0.97	0.98

[a]The predicted outflow profiles for SUCROSE co-administered with PRAZOSIN and PROPRANOLOL, obtained using these FCMA-predicted parameter values, are provided in Fig. S1 in the Data Supplement along with the predicted sucrose outflow profiles (Fig. S1) when SUCROSE was co-administered with PRAZOSIN and PROPRANOLOL.

§ The three predicted values of these four prazosin parameters are within 15% of each other.

∗ The three predicted values of these five propranolol parameters are within 15% of each other.

In Silico Liver

Drug Dose

Objects Representing Drug & Sucrose

Lobule

Zone 1 Zone 2 Zone 3 CV

Portal Vein Tract

Vascular Septum

Sinusoid Network

Portal Vein Tract

Zone 1

Zone 2

Zone 3

CV

Sinusoid Segment (SS)

Grid C
Grid B
Grid A
Core

SUCROSE

DRUG

CYP Binder

Hepatocyte

Figure 3

Figure 5

Prazosin

Propranolol