Supplemental Material can be found at:
http://dmd.aspetjournals.org/cgi/content/full/dmd.107.019067/DC1

0090-9556/08/3604-759–768$20.00
DRUG METABOLISM AND DISPOSITION
Vol. 36, No. 4
Copyright © 2008 by The American Society for Pharmacology and Experimental Therapeutics
19067/3323256
DMD 36:759–768, 2008
Printed in U.S.A.

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

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, California (L.Y., S.S.-B., C.A.H.); Department of Bioengineering and Therapeutic Sciences, the BioSystems Group, the University of California, San Francisco, California (S.P., G.E.P.R., C.A.H.); and Tempus Dictum, Inc., Eagle Creek, Oregon (G.E.P.R.)

Received October 28, 2007; accepted January 24, 2008

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 only sparse data available, all predicted disposition profiles were judged reasonable (within a factor of 2 of referent profile data). The order of precision, based on a similarity measure, was 3 > 2 > 1.

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 analogs 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. (2008), the simulated hepatic disposition of atenolol, antipyrine, labetalol, and diltiazem, along with coadministered 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 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] 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, physiologically based PK model parameters necessarily conflate features and properties of the biology (aspects of histology and others) with drug PCPs. In doing so, there is a risk that "the causal basis" becomes obscured because of 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 and rules). Our expectation has been that at some

This work was supported in part by grants (to C.A.H.) and fellowships (to S.S.-B. and S.P.) provided by the CDH Research Foundation and abstracted in part from the Ph.D. dissertation presented by L.Y. to the Graduate Division, University of California, Berkeley, CA.

Article, publication date, and citation information can be found at http://dmd.aspetjournals.org.

doi:10.1124/dmd.107.019067.

[S] The online version of this article (available at http://dmd.aspetjournals.org) contains supplemental material.

ABBREVIATIONS: ISL, in silico liver; PCP, physicochemical property; PK, pharmacokinetic; CV, central vein; SS, sinusoidal segment; PV, parameter value; CC, correlation coefficient; FCMA, fuzzy *c*-means algorithm; ANN, artificial neural network; SM, similarity measure.

Downloaded from dmd.aspetjournals.org by on June 12, 2009

DMD DRUG METABOLISM AND DISPOSITION

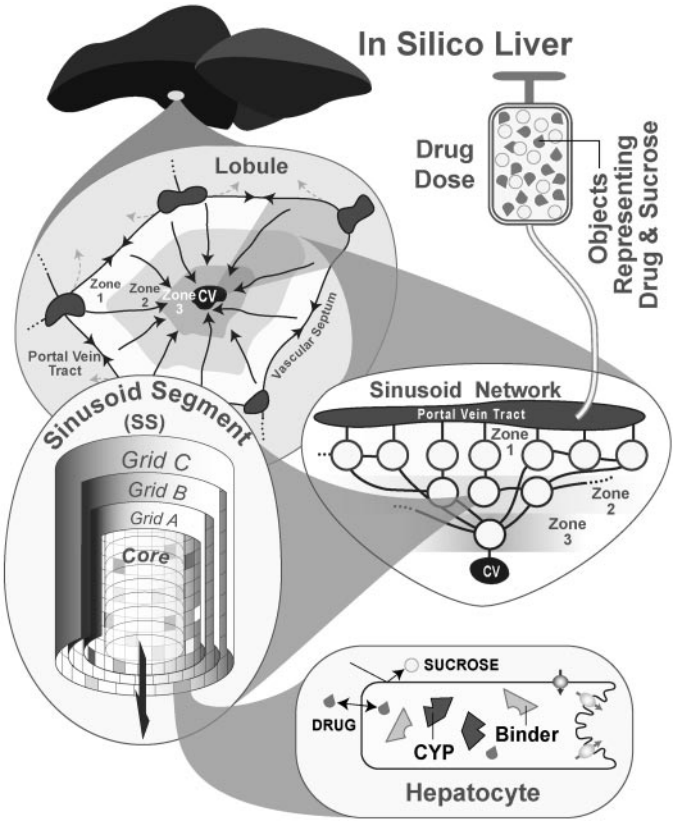

FIG. 1. Illustrations of ISL model components. The six levels of ISL organization, from largest to smallest, are 1) the ISL experiment system, 2) the SINUSOID network, 3) the SINUSOID segment, 4) SINUSOID spaces, 5) CELLS and EXTRACELLULAR domains, and 6) ENZYMES and binders. Upper left, a schematic diagram 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 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 SSs are placed at each graph node: they simulate sinusoid functions. Lower left, a schematic diagram 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, and grid C represents all other spaces, including the space of Dissé and hepatocytes. Lower right, a HEPATOCYTE container: objects 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).

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

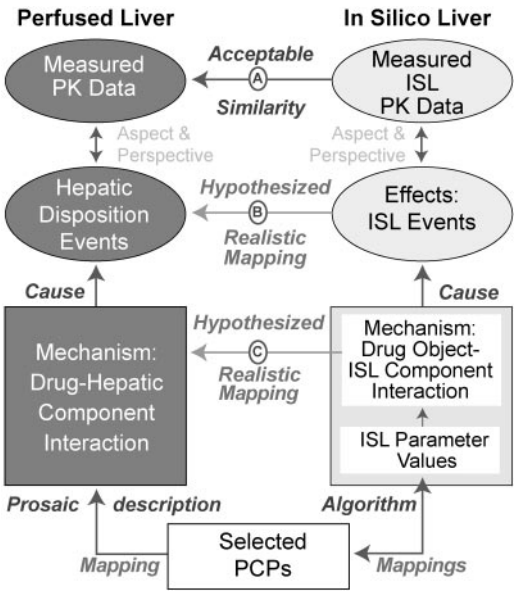

FIG. 2. Illustration of relationships between ISL mechanisms and its components, along with relationships between the ISL and referent perfused liver counterparts. To the extent that mappings A, B, and C are realistic, the relationship between a drug's PCPs and specific hepatic disposition events will 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.

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 10 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.

Materials and Methods

To distinguish clearly in silico compounds and processes from corresponding hepatic structures and processes, we use SMALL CAPS when referring to the former.

Downloaded from dmd.aspetjournals.org by on June 12, 2009

aspet

DMD DRUG METABOLISM AND DISPOSITION aspet

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 spaces	*A2BJumpProb*	0.1	0.1	0.35	0.9
	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 components	*BindersPerCellMin*	5	5	5	10
	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 coadministered with DILTIAZEM (compensates for different *ISL2WetLabScaling* values).

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. (2008) 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. (2008) along with all other ISL features. For convenience, they are also listed in Supplemental Table S1. Only 10 parameters were modified during the course of this study. Except as noted below, all other parameter values were the same as those specified in Supplemental 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 two-dimensional 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 with 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 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 (Supplemental 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 Supplemental 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 the 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 grid A, or jump 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. (2008).

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, 2006) and 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 (Supplemental Table S1) that specifically influence disposition events, a major simulation task in Yan et al. (2008) 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 10 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

Downloaded from dmd.aspetjournals.org by on June 12, 2009

DMD DRUG METABOLISM AND DISPOSITION aspet

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. (2008). Five of the 10 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.

Because of 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 ISL outflow profiles and their in situ counterparts (A); ISL events and hepatic disposition events (B); and local ISL mechanisms and their drug-hepatic component counterparts (C). 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 on the basis of 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 by 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 for 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 one is attempting to identify and unravel complicated mappings, more data are 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 subsystem. 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 encountered previously, 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). To validate against the expanded set of targeted attributes, a current mechanistic ISL feature may need to be replaced by a revised and 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: overfitting 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. (2008)]. At best, an ISL can be used to anticipate the behavior of a new compound in the context of considerable uncertainty, given its assumptions. As 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 are a rough approximation of what might be expected if all assumptions are valid. Furthermore, 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 with an 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 that 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 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. (2008). 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. In the following, 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-Means Clustering. To handle uncertainty that arises during mappings of this type, we used the fuzzy c-means algorithm (FCMA) introduced by Bezdek et al. (1984) and Pal and Bezdek (1995). It can be used to transform descriptive mappings into quantitative relationships. It starts with

Downloaded from dmd.aspetjournals.org by on June 12, 2009

DMD DRUG METABOLISM AND DISPOSITION aspet

TABLE 2

Physicochemical properties of sucrose and six cationic drugs

Category	Parameters	Sucrose	Atenolol	Antipyrine	Prazosin	Labetalol	Propranolol	Diltiazem
I. Size	Mol. wt.	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	$\log P_{app}$		0.14	0.33	1.88	2.69	3.10	3.53
III. Binding	f_{uB}		0.74	0.60	0.54	0.52	0.45	0.28
IV. Topology related	RB count	5	8	1	4	8	6	7
	TPSA	189.5	84.6	26.9	107.0	95.6	41.5	59.1
V. Hydrogen bound related	HBD count	8	3	0	1	4	2	0
	HBA count	11	4	2	8	4	3	5
	Tautomer count	1	2	0	3	7	0	0

$\log P_{app}$, log of the octanol-water partition coefficient at pH 7.4, 37°C; f_{uB}, unbound fraction of each drugs in the liver perfusion medium as determined in Hung et al. (2001); RB, number of rotatable bonds, a measure of molecular flexibility; amine C–N bonds are not considered because of their high rotational energy barrier; TPSA, topological polar surface area is a sum of fragment contributions; *O*- and *N*-centered polar fragments are considered (Ertl et al., 2000); HBD, hydrogen bond donor; HBA, hydrogen bond acceptor; tautomers are organic compounds that are interconvertible by tautomerization.

influence relationships between subsets of PCPs and subsets of PCP-sensitive, ISL PVs from validated ISLs.

Sheihk-Bahaei and Hunt (2006) used the FCMA and only limited drug-specific data to predict parameter values for their analog 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 that 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 to measured values. They established a single mapping between 13 selected PCPs and 14 PCP-sensitive model PVs (the ISL has just 10). 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. (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. Furthermore, the artificial neural network method discussed below provides a mapping between all 9 PCPs and all 10 PCP-sensitive, ISL 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 by 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 10 PCP-sensitive ISL parameters in Table 1 into four subgroups: those that reflect the influence of a COMPOUND'S 1) effective size, 2) movement between spaces, 3) binding to LOBULAR components, and 4) METABOLISM.

We took the nine PCPs in Table 2 and divided them into five groups, as follows, according to their perceived similarity in influence. Group I includes molecular weight and volume. Group II is the partition coefficient ($\log P_{app}$). Group III is the unbound fraction (f_{uB}). Group IV is topological parameters: number of rotatable bonds and topological polar surface area. These two PCPs are known to be good descriptors for characterizing drug absorption, including

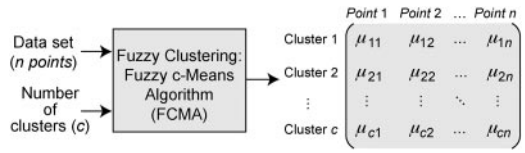

FIG. 3. Inputs and output of the FCMA. The FCMA locates c clusters within the set of n points (corresponding to n drugs) in PCP hyperspace and assigns a membership degree, μ_{ij}, for each point i to each cluster j.

intestinal absorption, bioavailability, Caco-2 permeability, and blood-brain barrier penetration (Ekins et al., 2000; van de Waterbeemd and Gifford, 2003). Group V is hydrogen bond-related parameters: hydrogen bond donor count, hydrogen bond acceptor 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 relationships between PCPs and ISL parameters are believed to be nonlinear, 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 (Fig. 4).
3. DRUG METABOLISM (ISL parameter group D) simulates a complicated process and so it may map to a broader range of PCPs. We included the following: $\log P_{app}$ (subgroup II) along with the PCPs in subgroups IV and V.
4. Both molecular weight 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

Downloaded from dmd.aspetjournals.org by on June 12, 2009

DMD

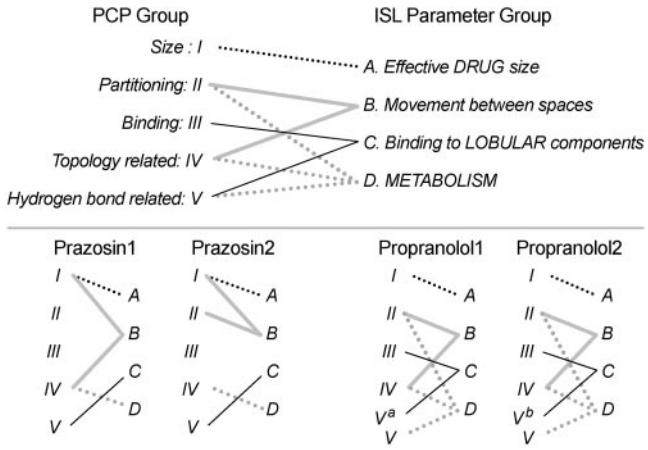

FIG. 4. Mappings between PCP subgroups and ISL parameter subgroups. The four ISL subgroups are specified in Table 1; the five PCP subgroups are specified in Table 2. Each line represents a hypothesized strong influence, e.g., the partitioning- and topology-related PCPs are thought to have a strong influence 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.

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 in 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 those 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 Supplemental Data.

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 $S_{\text{new}} = \{c_1, c_2 \ldots c_n, c_{n+1}\}$.

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.

TABLE 3

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

PCP subgroups II and IV: values comprising PCP subgroups II and IV. Cluster: the results of applying the FCMA to the five sets of three PCP values. ISL parameters: tuned *B2CJumpProb* values for the four drugs already studied along with the predicted value for prazosin.

Boldface: primary cluster membership.

	Atenolol	Antipyrine	Labetalol	Diltiazem	Prazosin
PCP subgroups II and IV					
RB count	8	1	8	7	4
TPSA	84.58	26.94	95.58	59.09	106.96
$\log P_{\text{app}}$	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)

RB, number of rotatable bonds; TPSA, topological polar surface area.

Else, let the G value be the number of group-mates of c_{n+1}.

Step 4. Call the q groups $G_1, G_2, \ldots G_q$, 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

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

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

$$\hat{\vec{P}}_{G_k} = \left(\sum_{j=1}^{m} \mu_j \cdot \vec{P}_j\right) / \sum \mu_j \tag{2}$$

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. 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 overfitted 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 utilized normalized values of the 4 sets of 9 PCPs in Table 2 as inputs; values of the 4 sets of 10 ISL PVs in Table 1 were obtained as outputs. We then used the trained ANN to separately predict a full set of 10 PCP-sensitive, ISL PVs for PRAZOSIN and PROPRANOLOL, given their full set of 9 PCPs as inputs.

Values were normalized as follows: divide each point by its vector Norm. For example, if drug (point) X has PCPs A, B, and C, then it was divided by its length N, i.e., $N^2 = \text{A}^2 + \text{B}^2 + \text{C}^2$ and $X_{\text{normalized}} = (\text{A}/N, \text{B}/N, \text{C}/N)$. 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 profiles with the lag-time adjusted, referent outflow profiles is detailed and discussed in Hunt et al. (2006) and Yan et al. (2008). Briefly, we assumed that the coefficients of variation of identically timed observations between $t = 1$ and $t = 100$ s from repeat experiments (referent) are constant. For each referent outflow profile measure P, we create two curves, $P^{\text{l}} = P(1 - d)$ and $P^{\text{u}} = P(1$

DRUG METABOLISM AND DISPOSITION

Downloaded from dmd.aspetjournals.org by on June 12, 2009

DMD DRUG METABOLISM AND DISPOSITION aspet

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 (boldface) was used to predict that ISL parameter value for PRAZOSIN and PROPRANOLOL, given their value of that PCP. Those 10 predicted parameters values are listed in Table 5 and were used to generate the outflow profiles in Fig. 5.

	ISL2WetLabScaling	*A2BJumpProb*	*B2AJumpProb*	*B2CJumpProb*	*C2BJumpProb*
Mol. wt.	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*
Mol. wt.	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
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

$+ d$). 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 S.D. 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. (2008), in which the ISL outflow profile was iteratively tuned to give a good match to the referent profile, we used $d = 1$ (S.D.), and an ISL outflow profile was deemed similar to the referent if 80% or more of ISL outflow values were within the band (SM $\geq$0.8). Here, using predicted PVs without any postprediction refinement, we speculated that having a predicted profile fall within a factor of 2 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 (http://oscar.openclustergroup.org/) running RedHat's Fedora 5. The distribution of the runs uses MPICH 1.2.7 (http://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 (http://swarm.org/wiki/Main_Page). The initial pseudorandom number 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) (http://www.r-project.org/) scripts, and Microsoft Excel. As done previously (Hunt et al., 2006; Yan et al., 2008), 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-s intervals was used.

Results

Predicted PCP-Sensitive ISL Parameter Values. Ninety correlations were performed: 4 values of each of 10 ISL PVs (from having already tuned the ISL for four drugs) with the 4 values of each of 9 PCPs corresponding to those four drugs. Table 4 lists the CC values. The regression line for the 10 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

TABLE 5

Predicted ISL parameter values for SUCROSE, PRAZOSIN, and PROPRANOLOL

For each of the three methods, the parameter values listed were combined with the PCP-insensitive parameter values. The resulting ISLs were used to generate the expected outflow profiles in Figs. 5 to 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[b]	0.22	0.40	0.21[b]	0.29	0.21[b]
B2CJumpProb	0.42[c]	0.50[b]	0.49	0.41[c]	0.50[b]	0.50[c]	0.50[b]
C2BJumpProb	0.56[c]	0.31	0.50	0.48[c]	0.22	0.42[c]	0.32
MetabolizeProb	0.34	0.13[b]		0.25	0.12[b]	0.19	0.09[b]
BindersPerCellMin	6	12		7	6	8	8
BindersPerCellMax	7	14		14	13	15	17
SoluteBindingProb	0.47[c]	0.39[b]		0.41[c]	0.42[b]	0.43[c]	0.42[b]
SoluteBindingCycles	24[c]	23[b]		23[c]	24[b]	24[c]	23[b]
SM values (±1 S.D.)	0.36	0.68		0.50	0.58	0.79	0.97
SM values (±1.5 S.D.)	0.55	0.82		0.77	0.72	0.97	0.98

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

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

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

Downloaded from dmd.aspetjournals.org by on June 12, 2009

DMD DRUG METABOLISM AND DISPOSITION

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 coadministered sucrose, eqs. 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 Supplemental Data along with the predicted sucrose outflow profiles (Supplemental Fig. S1) when SUCROSE was coadministered 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 (Supplemental 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 through 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., 2008), 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 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 *Materials and 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 S.D. flanking the original wet-lab outflow profile values as a measure of the similarity of ISL simulations to

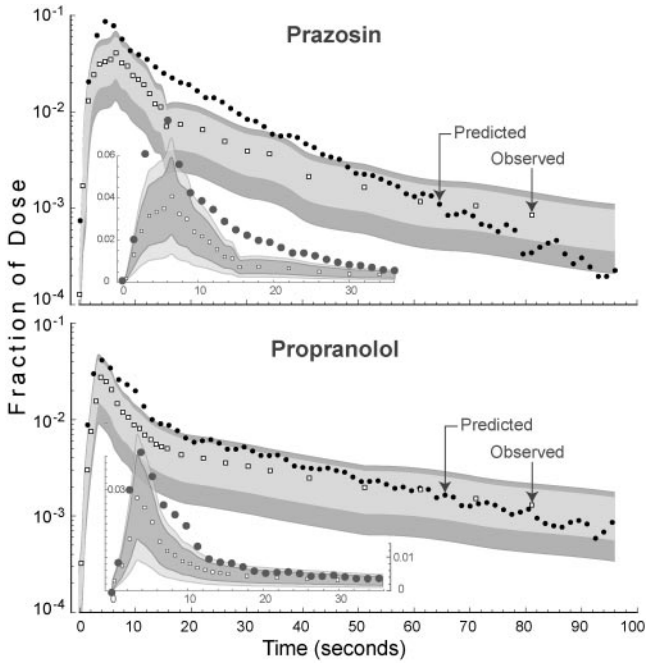

FIG. 5. Expected outflow profiles (●) for prazosin and propranolol using predicted parameter values. These semilog plots show results of simulation experiments for which the PCP-sensitive PVs were predicted using the simple linear regression method values listed in Table 5. All other ISL parameter values were identical to those reported in Yan et al. (2008). □, observed, referent wet-lab data. The data are the fraction of dose per outflow unit (per milliliter for the referent) as a function of time (1 simulation time unit = 1 s) after dosing with DRUG. The light gray band spans the range for the mean of the referent data ±1 S.D. as specified in the text. The dark gray band spans the range for the mean of the referent data ±1.5 S.D. Each ISL datum is the smoothed (window size: three values) mean value of 48 independent ISL runs.

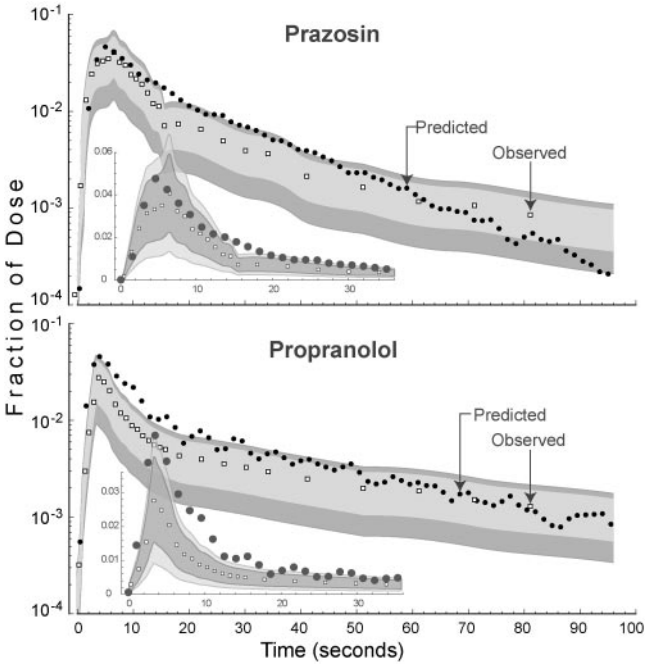

FIG. 6. Expected outflow profiles (●) for prazosin and propranolol using predicted parameter values. These semilog plots show the results of simulation experiments for which the PCP-sensitive PVs were predicted using the FCMA values listed in Table 5. Otherwise, the graph components are same as those specified in Fig. 5.

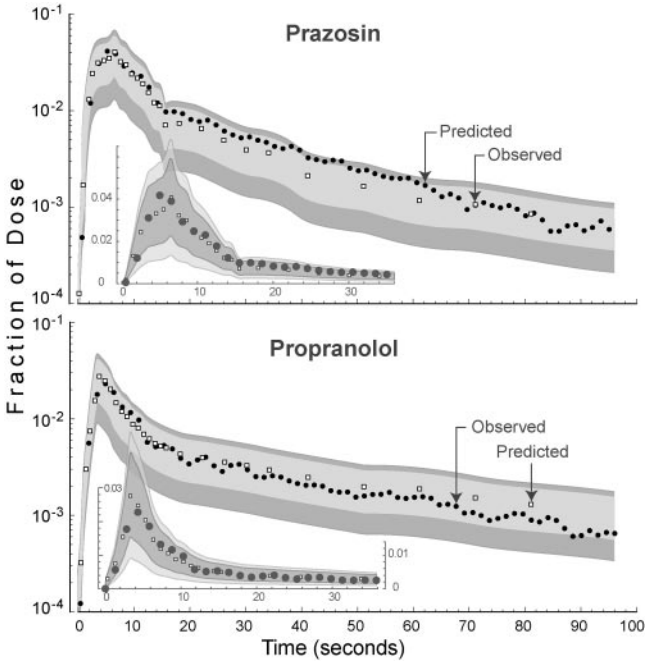

FIG. 7. Expected outflow profiles (●) for prazosin and propranolol using predicted parameter values. These semilog plots show results of simulation experiments for which the PCP-sensitive PVs were predicted using the ANN method values listed in Table 5. Otherwise, the graph components are same as those specified in the legend to Fig. 5.

Downloaded from dmd.aspetjournals.org by on June 12, 2009

DMD DRUG METABOLISM AND DISPOSITION aspet

those wet-lab results. For the outflow profiles generated from the predicted parameter values, two reference bands were used: ±1 S.D. and ±1.5 S.D., and SM values were calculated for each. For these data, a band of ±1 S.D. corresponds to a constant 45% coefficient of variation, and a band of ±1.5 S.D. corresponds to a constant 67.5% coefficient of variation. The ±1 S.D. 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 S.D SM values were as follows: 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 (Supplemental Fig. S2) along with the corresponding simulated outflow profiles (Supplemental Fig. S3) are provided in the Supplemental Data.

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 nonlinear 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 used clustering, specifically fuzzy clustering, as one of the three methods because it is difficult to classify PCPs and the drugs they characterize into crisp, disjoint clusters. Conventional clustering produces 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 nonlinear 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 subset 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 to 7 as a plausible estimate of what one might expect after administration of prazosin and propranolol to a perfused liver. In the case of the ANN (Fig. 7), we recognized the risks associated with overparameterization (Hudson and Cohen, 1999). All three methods gave rise to predicted profiles that were within a factor of 2 of the actual profiles. We found it noteworthy that the simple, single PCP correlation method (Fig. 5) led to ISL outflow profiles that were close to those of FCMA and ANN.

A single point in discretized PCP space represents a 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 in which different compounds have similar ISL PVs. Note also that for both PRAZOSIN and PROPRANOLOL, there are cases for which 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.

Downloaded from dmd.aspetjournals.org by on June 12, 2009

DMD DRUG METABOLISM AND DISPOSITION aspet

The results suggest that when use of 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 finding 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 subregion 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, 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, wherein 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 of 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 the 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 physiologically based PK, it seems important to use both the synthetic method for chronicling modeling assumptions (heuristic value) and the inductive method to achieve a high-tolerance predictive value. We posit that, by having simplified, yet fine-grained, heuristic models, such as 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 analogs that will enable us to explore, digest, categorize, understand, and use the massive amounts of data being generated by "omic" technologies.

Acknowledgments. We thank our collaborator Michael S. Roberts for providing us with the original liver perfusion data; we also thank 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. *Comput Geosci* **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, New York.

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 epitheliomimetic 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 (1999) *System Identification: Theory for the User,* 2nd ed, Prentice Hall PTR, Upper Saddle River, NJ.

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 *Proceedings of the 37th Conference on Winter Simulation* (Perrone LF, Lawson BG, Liu J, and Wieland FP eds); 2006 Dec 3–6; Monterey, CA, pp 1617–1624, Association for Computing Machinery, New York.

Sheikh-Bahaei S and Hunt CA (2007) In silico hepatocyte: agent-based modeling of the biliary excretion of drugs in vitro, in *Proceedings of the 2006 Agent-Directed Simulation Symposium* (Yilmaz L, et al. eds); 2006 April 3–9; Huntsville, AL, pp 147–163,SCS Press, San Diego, CA.

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 *Proceedings of the 20th International Conference on Computer Applications in Industry and Engineering* (Hu G ed); 2007 Nov 7–9; San Francisco, CA, pp 23–27, ISCA, Cary, NC.

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,* in press.

Address correspondence to: Dr. C. Anthony Hunt, Department of Bioengineering and Therapeutic Sciences, University of California, 513 Parnassus Ave., S-926, San Francisco, CA 94143. E-mail: a.hunt@ucsf.edu

Downloaded from dmd.aspetjournals.org by on June 12, 2009