

PREDICTION OF *IN VITRO* HEPATIC BILIARY EXCRETION USING STOCHASTIC AGENT-BASED MODELING AND FUZZY CLUSTERING

Shahab Sheikh-Bahaei¹

¹Joint Graduate Group in Bioengineering
University of California

Berkeley and San Francisco, CA 94143-0446 U.S.A.

C. Anthony Hunt^{1,2}

²Department of Biopharmaceutical Sciences
University of California

San Francisco, CA 94143-0446 U.S.A

ABSTRACT

We present a method for estimating (predicting) parameter values for an agent-based model of *in silico* hepatocytes (ISH). The method enables the ISH to interact with simulated drugs to reasonably match results from *in vitro* hepatocyte excretion studies. Further, we make the estimation method available to the model, itself, to enable it to reasonably anticipate (predict) the biliary transport and excretion properties of a new compound based on the acceptable parameter values for previously encountered compounds. We use Fuzzy c-Means (FCM) classification algorithm to determine the degree of similarity between previously tuned compounds and the new compound. Specifically, a set of simulation parameters for enkephalin was predicted using the tuned parameter values of salicylate, taurocholate, and methotrexate. The feature space for the FCM classification is the physicochemical properties of the compounds.

1 INTRODUCTION

Accurate estimation of a drug's hepatic disposition (including hepatic metabolism, protein binding, intracellular sequestration, and biliary excretion) is a crucial step in the development of clinically practical drugs. Early in the new drug development process there is often a considerable over supply of candidates. Suitable *in silico* methods can help narrow the list of candidate compounds to a manageable number before beginning expensive wet-lab evaluation and screening of those selected (Augen 2002, Rajasethupathy 2005). Compounds that are likely to have undesirable metabolic and excretory properties need to be culled from the list. The current most widely used *in silico* methods rely on correlational techniques rather than estimates based on knowledge of the mechanisms involved. We focus on modeling the biliary excretion of compounds at the mechanistic level. Biliary excretion is a relatively complex process involving translocation across the sinu-

soidal membrane, movement through the cytoplasm, and transport across the canalicular membrane. Different transporters and metabolic enzymes can be involved. Intracellular spatial organization may be important. Competitive interactions can occur between these components and other compounds undergoing the same processes.

Numerous *in vitro* systems (e.g., isolated perfused livers, isolated hepatocyte, short-term cultured hepatocyte couplets and long term sandwich cultured hepatocytes) have been used to investigate biliary excretion. In the *in vitro* sandwich-culture system, bile can be sequestered in spaces (small lumens) created by adjacent hepatocytes that have formed tight junctions between themselves, as illustrated in Figure 1A. The tight junctions form a seal between the luminal contents and external media. For the low resolution model described herein, the system has been broadly subdivided into three spaces: intracellular (cytosol), canalicular lumen, and the incubation medium. In the *in vitro* system, Ca^{2+} is responsible for maintaining the barrier function of the tight junctions that form a seal between the canalicular lumen and the incubation buffer. The barrier can be disrupted by depletion of Ca^{2+} . When that is done, the solution (biliary secretions from hepatocytes) that had accumulated in the canalicular lumen spaces mixes with the incubation medium. The cumulative uptake when the standard media is used represents the amount of substrate in both intracellular (cytosol) and in the canalicular lumen. However, when Ca-free media is used, the cumulative uptake represents the amount of substrate in cytosolic compartment only (Liu et al. 1999a). Thus, the amount of substrate excreted in the canalicular lumen (i.e., *in vitro* counterpart to biliary excretion) can be estimated from the difference between the cumulative uptake in presence and absence of Ca^{2+} . We constructed a low-resolution (few components; limited detail) model to test hypotheses about the mechanistic details of biliary transport *in vitro* and to predict transport and excretion properties of newly encountered compounds. The biliary excretion estimated by this method for several compounds is consistent with *in vivo* biliary excretion data (Liu et al. 1999a).

2 THE MODEL

We use agent-directed programming to construct the model. Within the *in silico* hepatocytes (ISH) we represent cells using fixed agents placed in a 2D grid where mobile objects representing solute can interact with them stochastically. The ISH used here is the same as that presented and described previously (Sheikh-Bahaei et al. 2006). The descriptions that follow are consequently brief, yet sufficient for clarity. Consult (Sheikh-Bahaei et al. 2006) for additional detail.

To avoid confusion hereafter and clearly distinguish *in vitro* components and features from their *in silico* counterparts, such as a “hepatocyte,” a “solute,” or “excreted,” we use SMALL CAPS when referring to the *in silico* system. We model the canalicular lumen (center, Figure 1A) as an object acting as a container inside the simulated HEPATOCYTE into which SOLUTES can be EXCRETED by TRANSPORTERS. Simply eliminating this space simulates Ca^{2+} disruption of tight junctions and mixing of what would have been luminal contents with the extracellular media, as illustrated in Figure 1. The sketch illustrates hepatocytes in a sandwich culture *in vitro* and the organization of the components within the ISH. Part A shows two hepatic cells (white), attached by tight junctions. A canalicular lumen space (shaded) is between them. The external medium includes Ca^{2+} . Part B shows the same system, but the Ca^{2+} in the media has been depleted breaking the tight junctions.

The Incubation Medium is represented by a 2D square grid in which HEPATOCYTES and SOLUTES can be placed to interact with each other. *Drug Compounds* (SOLUTES) are represented as independent, mobile objects that move around stochastically (using a Moore neighborhood), governed by the flow of the incubation medium. During an experiment, the event histories of SOLUTES (and other objects), such as SOLUTE that has been TRANSPORTED out of a CELL, or that has diffused into a CELL, can be tracked individually or as groups. SOLUTES are initially placed uniformly and randomly in the space external to HEPATOCYTES. At each simulation cycle, a SOLUTE may stay in place or move randomly in one of eight directions (N, NE, E, SE, S, SW, W or NW with a probability of 1/9). A SOLUTE may, depending on its properties, PARTITION into an encountered HEPATOCYTE. There is also a chance that it may be transported (actively imported) into the CELL by TRANSPORTERS. The pseudorandom numbers were generated from a uniform distribution using a Mersenne Twister random number generator.

HEPATOCYTES are constructed from objects that represent corresponding hepatocyte components and the environment: factors that can bind drug, enzymes, transporters, and a space for biliary excretion. A *Binder* is an object

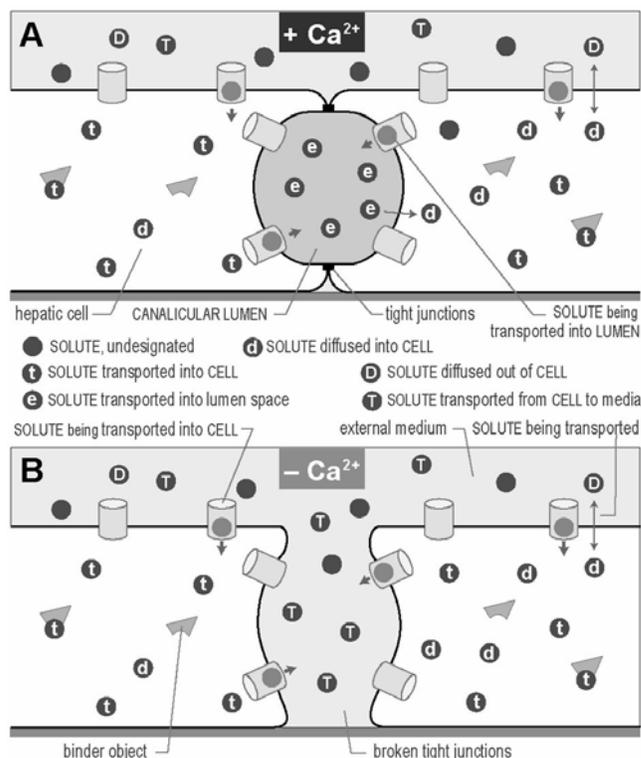


Figure 1: Two Adjacent Hepatic Cells in Standard (A) and Ca-free Buffers (B).

that can bind or sequester free SOLUTE and hold onto it for a specified number of binding cycles. An ENZYME is a specialized form of binder. It can “metabolize” a bound SOLUTE by replacing it, following the binding period, with a metabolite object and then destroying the replaced SOLUTE (for more details see Sheikh-Bahaei et al. 2005, Liu and Hunt 2005, and Liu and Hunt 2006). TRANSPORTERS belong to a subclass of binders. They can bind a free SOLUTE that is either inside or outside a CELL, and transport it to the opposite side of the CELL MEMBRANE, independent of the local SOLUTE concentration. Three important TRANSPORTER parameters are *Transport_in/out_probability* (it specifies the probability of binding a given SOLUTE), *Binding_cycles* (specifies how many cycles a SOLUTE remains attached), and *Excretion_space* (the location of excreted SOLUTES until removal to an EXTRACELLULAR space). In cultures with Ca^{2+} , SOLUTES in the *Excretion_space* are not removed simulating that they are “sealed” within a luminal space (e in Figure 1A). *Excretion_Mean* determines the average number of SOLUTE in this space.

2.1 In Silico Dynamics

Two parameters, *Solute_Membrane_Cross-In_Probability* and *Average_Cell_Capacity*, determine when a free SOLUTE in the “incubation medium” space that has encoun-

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

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

Once an intracellular solute is bound, there is chance that it will get excreted based on an exponential probability distribution with mean *Excrete_Mean*. If excreted, it will be removed by external spaces. If not removed, the list represents solute “sealed” between hepatocytes. If not excreted, the solute is treated as if it had been transported out by a basolateral membrane transporter, and is transferred back to the 2D extracellular space (the simulated incubation medium) in the next simulation cycle. SOLUTE within HEPATOCYTES maps to substrate in the cytosol. Solute in the *excretion space* maps to material excreted into the canalicular lumen (bile). For both standard and Ca-free media, the *in silico* cumulative uptake was calculated as:

$$Uptake_{in\ silico} = total\ of\ (partitioned-in + transported-in + excreted)\ solute \quad (1)$$

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

2.2 Parameter Tuning

The ISH parameter space consists of several different parameter types, some of which directly map to measurable biological counterparts. However, most of the parameters are simulation-specific with no direct biological significance. The ISH behavior space partially overlaps with the behaviors of the referent system when parameters have been appropriately tuned (adjusted). By so doing, we make the ISH behavior space converge on the behavior space of the referent system (Figure 3). In general, the parameter hyperspace of this model consists of both biologically meaningful (e.g., cell/compound density) and simulation-specific parameters (e.g., binding or membrane-crossing probabilities). Whereas the simulation-specific parameters have no particular constraints, the biologically meaningful parameters are constrained to stay within reasonable values (so that, at a minimum, outcomes are consistent with cell life). Optimizing the whole parameter vector makes it eas-

ier to find solutions within the constrained regions. A failure of optimization might be caused by false (biologically unrealistic) mechanisms implemented in the model. On the other hand, a successful optimization provides some measure of validation for the implemented mechanisms. Doing so, however, is not a goal of this paper: we have already validated this model against data for four compounds (Sheikh-Bahaei et al. 2006).

Parameter tuning can be done by optimization methods. Because the model is stochastic and therefore has discontinuity and noise in its behavior space, ordinary gradient-based Newton and quasi-Newton optimization techniques, which are commonly used for differential equations parameter tuning, cannot be applied. Non-gradient-based methods such as GAs (genetic algorithms) or Nelder and Mead Simplex method (Nelder and Mead 1965) are more successful.

After each simulation experiment a similarity measure (SM) algorithm (Ropella and Hunt 2003) assigns a score to the output of the simulation. This score provides a measure of similarity of the current output to the referent data. The goal is to maximize the SM score. The optimization algorithm used is the Nelder and Mead Simplex (Nelder and Mead 1965) method that has been widely used (Luersen et al. 2002, Tan et al. 2003, Chelouah et al. 2003, Lagariaset et al. 1998) to solve parameter estimation problems with discontinuities or where function values are noisy. Among the several different versions and extensions of this method, we are using the one described in (Neddermeijer et al. 2000) with minor changes.

2.3 Fuzzy C-Means Algorithm

Since Fuzzy Set Theory (FST) began in the 60's it has been developed as an alternative to probability theory in modeling uncertainty. Pattern recognition, or search for structure in data, provided the early motivation for developing FST because of the fundamental involvement of human perception (Derrig and Ostaszewski 1995).

A fuzzy classifier provides a measure of the degree to which a pattern fits within a class. There are several techniques for fuzzy pattern recognition. A Fuzzy classifier based on a genetic algorithm requires a long training time and that time can increase dramatically when the training data has a high dimension. A fuzzy classifier having an ellipsoidal region has performed well in many classification problems, but it needs a vast amount of time to calculate the covariance matrix of the ellipsoidal clusters. K-Means-Based fuzzy classifier (Wong et al. 2000) uses the K-means algorithm to partition the training data for each class into several clusters, and then some fuzzy rules are used to construct a fuzzy classifier. In this work, we use a fuzzy pattern recognition technique introduced by Bezdek and described in (Bezdek et al. 1984): Fuzzy C-Means (FCM) iterative algorithm.

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

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

where there are n observations (in our case the physicochemical properties of the compounds of interest), c classes, μ_{ik} is the membership of x_k (here x is the vector of physicochemical properties) in class i , m (the “fuzzy exponent”) is a parameter used to control the fuzziness of the class allocation, and

$$d_{ik}^2 = (x_k - v_i)^T G (x_k - v_i) \quad (3)$$

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

$$v_i = \frac{\sum_{k=1}^n \mu_{ik}^m x_k}{\sum_{k=1}^n \mu_{ik}^m} \quad (4)$$

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

2.4 Parameter Estimation

In this section we present an algorithm which uses FCM to estimate the simulation parameters of enkephalin knowing the tuned parameters of salicylate, taurocholate, and methotrexate. Three compounds are minimal. It is, however, sufficient to demonstrate the approach. In an industry setting data on many more compounds will be available. The four compounds were classified to two and three clusters

using Fuzzy c-Means algorithm based on selections of physicochemical properties that are expected to contribute to biliary clearance of the compounds. Initially the following physicochemical properties were considered: molecular weight, logP, hydrogen bond donor count, hydrogen bond acceptor count, rotatable bond count, tautomer count, pKa, TPSA, volume, GPCR ligand, ion channel modulator, kinase inhibitor, and nuclear receptor ligand. The results are shown in Table 1. (For Table 2, property values were obtained from the following sources: <http://www.molinspiration.com/cgi-bin/properties>, http://www.syrres.com/esc/est_kowdemo.htm, and <http://ibmlc2.chem.uga.edu/sparc/index.cfm>); abbreviations are MW: molecular weight, HBD: hydrogen bond donor, HBA: hydrogen bond acceptor, RB: rotatable bond, TPSA: topological polar surface area, GPCR: G-protein-coupled receptor, NR: nuclear receptor, and IC: ion channel.)

Table 1: Fuzzy Classification Results of Salicylate, Taurocholate, Enkephalin and Methotrexate Based on Their Physicochemical Properties (Table 2), Where C is Number of Clusters.

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

Table 2: Physicochemical Properties of Salicylate, Taurocholate, Methotrexate and Enkephalin.

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

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

Knowing the similarity of the compounds in the physicochemical domain we approximated their model param-

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

In general, for a data set, *S*, containing *n* compounds $S = \{c_1, c_2, \dots, c_n\}$, the following algorithm is used to estimate the simulation parameters of a new compound, c_{n+1} :

- Step 1. Let $q = n$, and $S_{new} = \{c_1, c_2, \dots, c_n, c_{n+1}\}$.
- Step 2. If $q = 1$ quit. Else, classify S_{new} into q clusters using FCM.
- Step 3. If c_{n+1} is not in the same group with at least another compound, then decrease q to $q-1$. Repeat steps 2 and 3.
- Step 4. Let's call the q groups G_1, G_2, \dots, G_q where $c_{n+1} \in G_1$. Let μ_k be the membership degree of c_{n+1} to group k . Estimate the simulation parameters of c_{n+1} as:

$$\hat{P}_X = \sum_{k=1}^q \mu_k \cdot \bar{P}_{Gk} \quad (5)$$

where \bar{P}_{Gk} is the average parameter vector of all the members of group k .

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

3 RESULTS

The referent for this model is an *in vitro* system used for studying primary rat hepatocytes (Liu et al. 1999a). Liu et al. (1999a) show that hepatocytes cultured in a collagen-sandwich configuration for up to five days establish intact canalicular networks, and reestablish polarized excretion of organic anions and bile acids. The system is a useful *in vitro* model for investigating the hepatobiliary disposition of compounds. The authors report that after the cells have been maintained in sandwich culture for five days, the cumulative uptake of [³H] taurocholate (a common component of bile) by the hepatocytes was significantly higher in standard Ca-containing media, compared with that of Ca-free media. The difference is a consequence of accumulation of taurocholate in intracellular canalicular spaces established during the prior culture period. [³H]

Taurocholate efflux from cells pre-loaded with drug for five days was greater in Ca-free compared with standard Ca-containing media. There is, of course, variability both within and between experiments.

Figure 2 shows the correlations between *in vitro* hepatocyte and *in silico* ISH uptake values at different times. The *in vitro* values were obtained (Liu et al. 1999a, Liu et al. 1999b) using well-established, sandwich-cultured hepatocytes using both standard and Ca-free media. The *in silico* uptake values were calculated using Equation (1). The ISH parameter values for each drug were iteratively optimized as reported in (Sheikh-Bahaei et al. 2006).

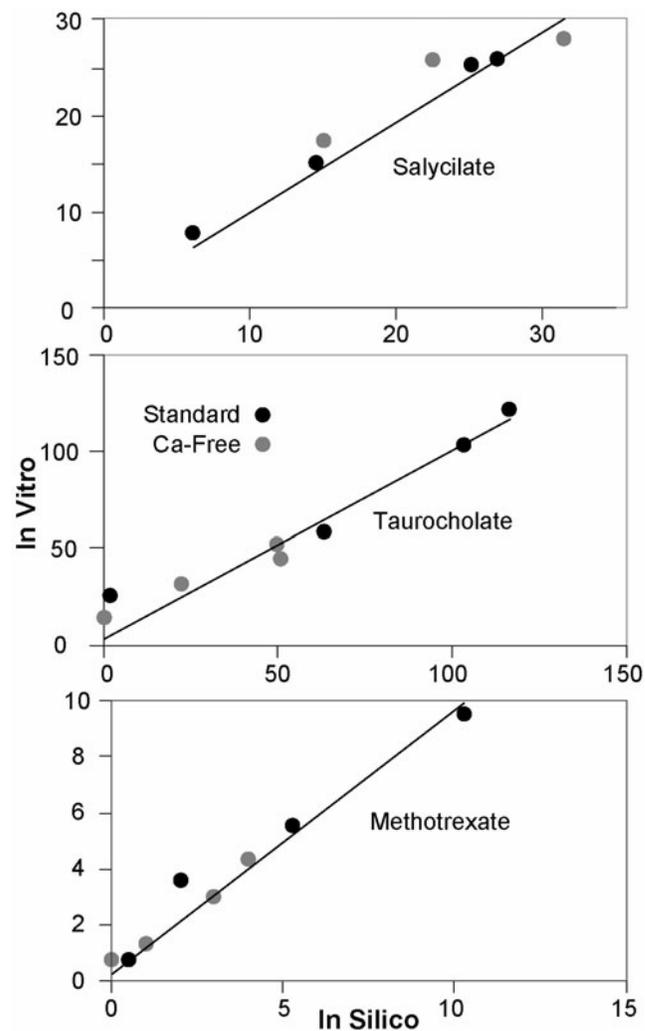


Figure 2: Correlation Between *In Vitro* and *In Silico* Uptake Values at Four Different Times (0.5, 2, 5 and 10 minutes) For Standard (Black Circles) and Ca-free Media (Gray Circles).

The measured uptake of enkephalin in well-established, sandwich-cultured hepatocytes, using both standard and Ca-free media, is presented in Figure 4A. Also shown is the model-predicted time course of *in vitro*

enkephalin uptake under those same two conditions. In Figure 3 the matches to both types of uptake are shown when model parameter values are iteratively tuned (Sheikh-Bahaei et al. 2006). Note that the *in vitro* uptake of enkephalin is about four times that of taurocholate, which has the largest uptake of the three compounds. The predicted uptake of enkephalin in the Ca-free buffer reaches its steady state too early. This might be due to the saturation of the uptake TRANSPORTERS used in the model. The *in silico* concentration of enkephalin is much higher than for the other three, and that may be a reason for saturation.

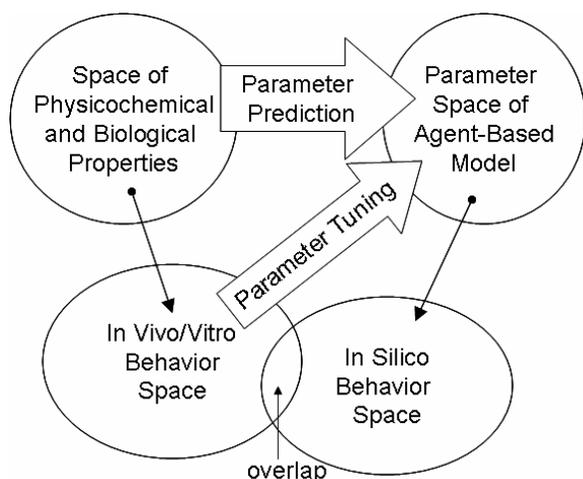


Figure 3: The Relationship Between In Silico Parameter/Behavior Spaces and Their In Vitro Physicochemical/Behavior Spaces.

4 CONCLUSION AND DISCUSSION

In Sheikh-Bahaei et al. (2005, 2006) we present an earlier version of the ISH which is, upon properly tuning its parameters, capable of mimicking its biological referents. The ISH parameter space consists of both biologically meaningful and simulation-specific parameters. The ISH behavior space partially overlaps that of the referent system. By properly adjusting parameters, we make the ISH behavior space converge on the behavior space of the referent system. In this paper, we use the ISH to *predict* the behavior of the referent system when introduced to a new compound not previously encountered. In order to predict that behavior, we needed means to *estimate* (predict) an appropriate set of parameter values. Note that parameter prediction is a direct mapping from physicochemical properties to the ISH parameter space whereas parameter tuning draws its information from the biological behavior space (Figure 3). In this work we presented an algorithm (Section 2.4) to estimate those values. It utilizes the FCM algorithm to cluster previously encountered compounds

based on their physicochemical properties (Figure 3). Our FCM approach offers three important advantages:

1. Because FCM is an unsupervised learning method (to work, it does not need to be trained with training data) the problem of over-fitting to the training data is minimized. That is particularly important in our case due to the small size of our data set.

2. It provides soft clusters. They are more biologically realistic.

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

We now discuss the acceptability of the results in Figure 4. Within experiments, there can be orders of magnitude differences between active and passive uptake of different compounds along with comparable differences in biliary excretion. Typically, the uptake and biliary excretion values of the same compound, between experiments, using essentially the same experimental system, is within a factor of two. The results of simulations can be more precise, but without drawing on additional information, their predictive accuracy can not be better. The simulations described here are intended to be analogous to repeat *in vitro* experiments: an observer should not be able to distinguish between data coming from a repeat *in vitro* experiment and data from an experiment on a tuned ISH. With those model use expectations set, the results in Figure 4 are minimally acceptable. They are *good enough* to facilitate decision making during early drug selection and development. The simulated enkephalin results are based on the uptake data of only three other compounds. Our expectation is that when the above procedure is used with descendants of model in Figure 1 to predict the uptake and excretion properties of future new compounds, the usefulness of predicted results will improve with each expansion of the set of successfully represented compounds.

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

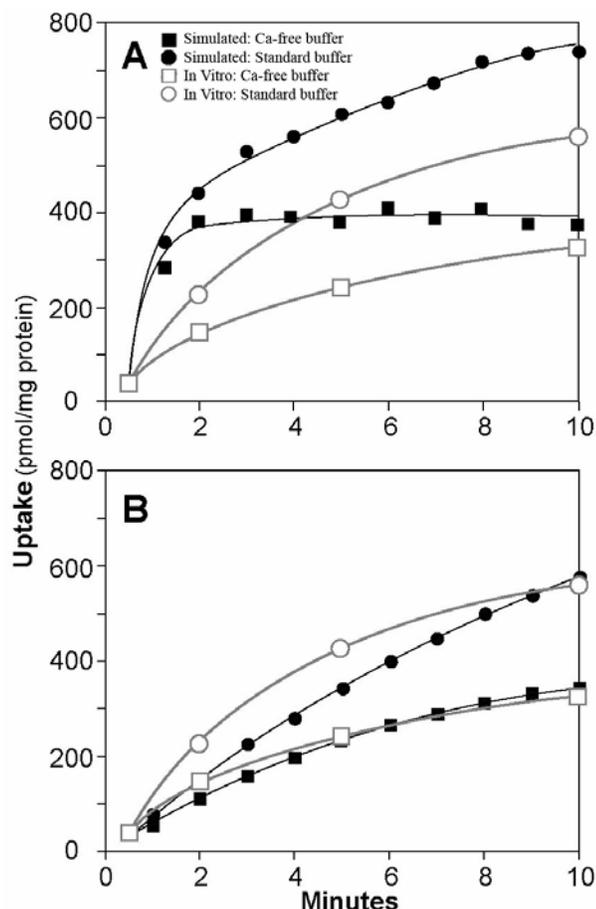


Figure 4: Comparisons of enkephalin uptake, *in vitro* and simulated, for two different growth conditions; the simulations use either the estimated (A) or tuned (B) parameter values (Table 3).

Table 3: Tuned Vs. Estimated Parameter Values for Enkephalin.

Parameter	Tuned	Estimated
artHepExcretionMean	1.20	0.51
artBindersPerCellMin	5	5
artBindersPerCellMax	10	10
artCellAverageCapacity	0.19	1.15
artTransportersPerCellMin	5	5
artTransportersPerCellMax	10	10
artSoluteTransportInProb	0.016	0.023
artSoluteTransportOutProb	0.0808	0.14
artSoluteTransportCycles	1	2
artSoluteMembraneCrossInProb	0.012	0.035
artSoluteMembraneCrossOutProb	0.095	0.37
artMetabolizationProb	0	0
artSoluteBindingProb	0.032	0.052
artSoluteBindingCycles	4	3

biological properties of compounds of interest. The approach is new: more work is needed to uncover and understand limitations and to delineate advantages relative to other methods typically used (optimization, data fitting etc.).

ACKNOWLEDGMENTS

This research was funded in part by the CDH Research Foundation (of which CAH is a Trustee) and CAH. We thank Nasim Sassan for assisting with evaluation of *in vitro* systems and data, Pearl Johnson and Maria Cabahug for assisting with manuscript preparation. We thank Glen Ropella and members of the BioSystems Group for helpful discussion and commentary.

REFERENCES

- Augen, J. 2002. The evolving role of information technology in the drug discovery process. *Drug Discovery Today* 7(5): 315-23.
- Bezdek, J. C., R. Ehrlich, and W. Full. 1984. FCM: The fuzzy c-means clustering algorithm. *Computers and Geoscience* 10: 191-203.
- Chelouah, R., and P. Siarry. 2003. Genetic and Nelder-Mead algorithms hybridized for a more accurate global optimization of continuous multimimima functions. *European Journal of Operational Research* 148: 335-348.
- Derrig, R. A., and K. M. Ostaszewski. 1995. Fuzzy techniques of pattern recognition in risk and claim classification. *Journal of Risk and Insurance* 62: 447-482.
- Hunt, C. A., G. E. P. Ropella, M. S. Roberts, and L. Yan. 2004. Biomimetic *in silico* devices. *Computational Methods in Systems Biology. Lecture Notes in Bioinformatics* 3082: 34-42. Available via <<http://biosystems.ucsf.edu/Research/RecentPapers.html>> [accessed April 8, 2006].
- Johns Hopkins Pathology. 2006. Gallbladder & bile duct cancer. *Anatomy and physiology of the gallbladder and bile ducts*. Available via <<http://pathology2.jhu.edu/gbbd/anatphys.cfm>> [accessed April 8, 2006].
- Liu, X., E. L. LeCluyse, K. R. Brouwer, L-S. L. Gan, LeMasters, J. J. Stieger, B. P. J. Meier, and K. L. R. Brouwer. 1999a. Biliary excretion in primary rat hepatocytes cultured in a collagen-sandwich configuration. *American Journal of Physiology* 277: G12-G21.
- Liu, X., J. P. Chism, E. L. LeCluyse, K. R. Brouwer, and K. L. R. Brouwer. 1999b. Correlation of biliary excretion in sandwich-cultured rat hepatocytes and *in vivo* in rats. *Drug Metabolism and Disposition* 27: 637-644.
- Liu, Y., and C.A. Hunt 2005. Studies of intestinal drug transport using an *in silico* epithelio-mimetic device. *Biosystems* 82(2): 154-167.

- Liu, Y., and Hunt, C. A. 2006. Mechanistic study of the interplay of intestinal transport and metabolism using the synthetic modeling method. *Pharmaceutical Research* 23(3): 493-505.
- Luersen, M. A., and R.L. Riche. 2002. Globalized Nelder-Mead method for engineering optimization. In *Proceedings of the Third International Conference on Engineering Computational Technology (Stirling, Scotland, Sept 4-6)*, 165-166. Civil-Comp Press, Edinburgh, Scotland.
- Lagarias, J. C., J. A. Reeds, M. H. Wright, and P. E. Wright. 1998. Convergence properties of the Nelder-Mead simplex method in low dimensions. *SIAM Journal on Optimization* 9(1):112-147.
- Nelder, J. A., and R. Mead. 1965. A simplex method for function minimization. *Computer Journal* 7:308-313.
- Neddermeijer, H. G., G. J. Oortmarssen, N. Piersma, and R. Dekker. 2000. Adaptive extensions of the Nelder and Mead simplex method for optimization of stochastic simulation models. *Econometric Institute Report* 199. Econometric Institute, Erasmus University, Rotterdam.
- Rajasethupathy P., S. J. Vayttaden, and U. S. Bhalla. 2005. Systems modeling: a pathway to drug discovery. *Current Opinion in Chemical Biology* 9(4):400-6.
- Ropella, G. E. P., and C. A. Hunt. 2003. Prerequisites for effective experimentation in computational biology. In *Proceedings of the 25th Annual International Conference of the Engineering in Medicine and Biology Society (Cancun, Sept 17-21)*, 1272-5.
- Ropella, G. E., C. A. Hunt, and D. A. Nag. 2005a. Using heuristic models to bridge the gap between analytic and experimental models in biology. In *Proceedings of the 2005 Agent-Directed Simulation Symposium (San Diego, CA, Apr 2-8)*, *Simulation Series vol. 37, no. 2*, ed. L. Yilmaz, SCS Press, San Diego, CA, 182-190.
- Ropella, G. E., C. A. Hunt, and S. Sheikh-Bahaei. 2005b. Methodological considerations of heuristic modeling of biological systems. *The 9th World Multi-Conference on Systemics, Cybernetics and Informatics (Orlando, FL, July 10-13)*.
- Ropella, G. E. P., D. A. Nag and C. A. Hunt. 2003. Similarity measures for automated comparison of in silico and in vitro experimental results. In *Proceedings of the 25th Annual International Conference of the Engineering in Medicine and Biology Society (Cancun, Sept 17-21)*, 2933-2936.
- Sheikh-Bahaei, S., G. E. P. Ropella, and C. A. Hunt. 2005. Agent-based simulation of in vitro hepatic drug metabolism: in silico hepatic intrinsic clearance. In *Proceedings of the 2005 Agent-Directed Simulation Symposium (San Diego, CA, Apr 2-8)*, *Simulation Series vol. 37(2)*, ed. L. Yilmaz, SCS Press, San Diego, CA, 171-6.
- Sheikh-Bahaei, S., G. E. P. Ropella, and C. A. Hunt. 2006. In silico hepatocyte: agent-based modeling of the biliary excretion of drugs in vitro. In *Proceedings of the 2006 Agent-Directed Simulation Symposium (Huntsville, AL, Apr 3-9)*, *Simulation Series*, ed. L. Yilmaz, et al., SCS Press, San Diego, CA (ISBN 1-56555-303-9), 157-163.
- Takahashi, K., K. Kaizu, B. Hu, and M. Tomita. 2004. A multi-algorithm, multi-timescale method for cell simulation. *Bioinformatics* 20(4):538-546.
- Tan, S. Y. G. L., G. J. Oortmarssen, and N. Piersma. 2003. Estimating parameters of a microsimulation model for breast cancer screening using the score function method. *Annals of Operations Research* 119(1-4): 43-61.
- Wong, C.-C., C.-C. Chen and S.-L. Yeh. 2000. K-means based fuzzy classifier design. *IEEE International Conference of Fuzzy Systems* 1:48-52.

AUTHOR BIOGRAPHIES

SHAHAB SHEIKH-BAHAEI is a Ph.D. candidate in the joint Bioengineering graduate program at the University of California, Berkeley and San Francisco. His interest areas include biological modeling and simulation, artificial intelligence, automation, robotics, and discrete event systems. His e-mail address is <shahabsb@berkeley.edu> and his web page is <<http://ist-socrates.berkeley.edu/~shahabsb>>.

C. ANTHONY HUNT has served as a Professor of Biopharmaceutical Sciences and Bioengineering at the University of California, San Francisco for over 20 years. He is a faculty member for the graduate programs in Bioengineering, Biological and Medical Informatics, and Pharmaceutical Sciences and Pharmacogenomics. Three of his areas of interest are systems biology, computational biology, and pharmaceutical sciences. His e-mail address is <a.hunt@ucsf.edu> and his web page is <<http://biosystems.ucsf.edu/People/CAHunt/CAH.htm>>