Multi-Agent Based Modeling of Liver Detoxification

Shahab Sheikh-Bahaei[1], Sean H. J. Kim[1], and C. Anthony Hunt[1,2]

[1] **Joint Graduate Group in Bioengineering, University of California, Berkeley and San Francisco, California**
[2] **Department of Bioengineering and Therapeutic Sciences, The BioSystems Group,**
University of California, 513 Parnassus Ave., San Francisco, California 94143-0912

shahabsb@berkeley.edu   seanhjk@berkeley.edu   a.hunt@ucsf.edu

Abstract

Using game theory and reinforcement learning, we created and analyzed generalized agent-based models of hepatic toxin elimination processes to explore plausible causes of hepatic functional zonation. We considered a general situation in which a group of protective agents (analogous to liver cells) cooperate and self-organize their efforts to minimize optimally the negative effects of toxin intrusions. The model suggests that in order to do so, the agents should adjust their resource consumption based on two factors: 1) their ranked proximity to the common wealth and 2) the potential damage caused by toxins. We verified that liver cells do the same.

1. INTRODUCTION

The liver performs a wide range of functions including detoxification of blood-born compounds. So doing protects the body. A human cannot live more than 24 hours without the liver. Hepatocytes, the parenchymal cells of the liver, are among the most complex cells in the body. They cooperate with each other to detoxify xenobiotics by metabolizing them to less toxic compounds. Over the course of their evolution they have learned to do so in an effective and optimal way. Hepatocytes express heterogeneous, location-dependent enzyme and transporter activities to facilitate detoxification, apparently following an intrinsic agenda, the principles of which are not fully understood. This phenomenon is known as liver zonation [1]. To gain insight into those processes, we constructed and analyzed a generalized problem of cooperative[1] agents protecting their common wealth from harmful intruders.

1.1 Biology Background

The liver is a complex biochemical factory which synthesizes, modifies and metabolizes thousands of substances daily and provides the body with essential substances such as proteins and fats. The liver is also responsible for eliminating toxins and drugs that find their way into blood. The rate of elimination, known as *hepatic clearance*, is different for each compound. Histologically, the liver is divided into lobules. Lobules consist of hepatocytes arranged in a roughly cylindrical or sphere shape. The central vein, through which blood exits, is at the center. At the periphery are portal triads. The function of the liver is often described as being organized into three zones: periportal (upstream or zone 1), which encircles the portal tracts where blood enters, middle (zone 2), and perivenous (downstream or zone 3), which is poorly oxygenated and located around central vein. Oxygenated blood enters upstream, passes through the midzone, and exits downstream. Because of this spatial topology, the liver cells do not have the same exposure to incoming resources and compounds. For example, nutrients (e.g. oxygen) are more available to the upstream zone than to the downstream zone.

Hepatocytes, although genetically identical, express heterogeneous enzyme and transporter activities depending on their location within the lobule. For example, under normal conditions, hepatocytes located downstream express more
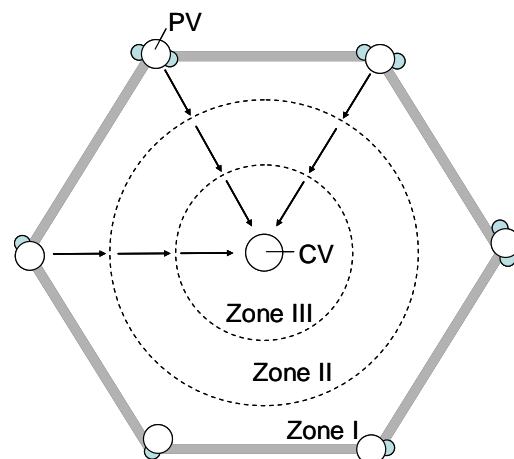

Figure 1. Cross section of a liver lobule. PV: portal vein, CV: central vein, arrows: direction of blood flow. Oxygenated blood enters the lobule from portal veins, and exits from central vein. Usually a liver lobule is described as being divided into three functional zones.

---

[1]Agents are assumed to have incomplete information about each other and cannot form coalitions; thus, the game does not fall into the class of Cooperative Games.

xenobiotic metabolism enzymes than do upstream hepatocytes. An obvious question is: why?

To represent hepatic metabolic zonation in physiologically based, pharmacokinetic models, researchers [2][3] usually divide the liver into compartments, each representing a different intrahepatic zone. Using a different approach, a "mechanistic model" is presented by [4] which proposes that zonation is induced by portocentral signal gradients.
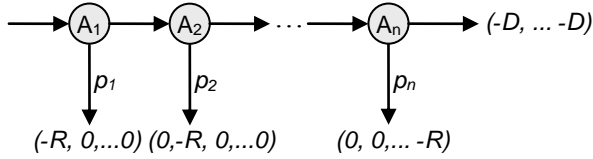

Figure 2. The game-theoretic model of the system. Agents either ignore or eliminate intruders. If Agent i (A_i) eliminates an intruder, it pays the cost of resource consumption (R). If all agents ignore an intruder, all must pay cost D, which is a consequence of damage caused by the intruder to the common wealth. p_i is the elimination strategy of agent i: it is the probability that A_i eliminates an incoming intruder. A_i's ability to eliminate is limited by $maxE_i$ $(0 \leq maxEi \leq 1)$.

2. METHODS

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

To understand the costs and benefits associated with liver zonation, we began by using the game theoretic model as shown in Fig. 2: a sequence of n agents is protecting their common wealth (all extrahepatic tissues) against intruders. Agents are limited in their ability to eliminate intruders. The goal of each is to minimize potential damage caused by the intruders while minimizing resource consumption. What is the optimal elimination strategy for agent i? Obviously, an optimal strategy for agent i depends on the strategies of other agents who have the same goal.

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

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

$$e_1 = 1 - maxE_1 \cdot p_1$$

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

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

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

Expected average costs of resource consumption are:

$$ARC_1 = maxE_1 \cdot p_1 \cdot R$$

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

Total expected costs are:

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

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

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

$$<Cost_2> = maxE_2 \cdot p_2 \cdot (1 - maxE_1 \cdot p_1)(R - D) + (1 - maxE_1 \cdot p_1) \cdot D \quad \text{(Eq.2)}$$

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

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

The vector field is shown in Fig. 3 for $D/R = 0.6$, 1.1, 1.7 and 2.3. It is easy to find the equilibrium of each game by inspecting its vector field. The figure shows that the equilibrium changes as D/R increases. When damage is small ($D/R < 1$), it is not cost effective to eliminate intruders. When damage is moderate, only the downstream agent expends elimination effort. When damage is large, both agents expend maximum elimination effort.

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

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

$$ADC = e_n \cdot D$$

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

$$<Cost_i> = ARC_i + ADC$$

$$<Cost_i> = e_{i-1} \cdot maxE_i \cdot p_i \cdot R + e_n \cdot D \qquad \text{(Eq.3)}$$

where e_i is the fraction intruders that escape from agent i; ADC is the average damage cost to each agent; ARC_i is the average cost of resource consumption to agent i; and $<Cost_i>$ is the total expected cost (due to both actions) to agent i. The vector field can be calculated the same as for the two-player game, but it is infeasible to visualize and find equilibria. In general, analyzing equilibria of games involving three or more players is hard [5][6].

The above analysis requires that all agents have a priori knowledge about other agents and the environment. All the actions available to other agents and all costs with all combinations of actions are known by all agents. Consequently, it does not offer a mechanism through which agents can reach an optimal strategy.

We considered a more realistic situation in which agents do not have a priori information about their environ-

ment (including other agents). We used multi-agent simulation and enabled agents to learn from experience following a simple reinforcement learning rule. By keeping track of accumulated reward (and penalty), agents could be reinforced to learn an optimal clearance strategy. Their task was to maximize the long-term average reward per action.

The *Q*-learning algorithm [7], a well known, reinforcement learning algorithm, has been shown to converge to an optimal decision policy. *Q*-learning has a solid foundation in the theory of Markov decision processes [8]. It is easy to implement and has been used widely in both single-agent and multi-agent contexts (see [8] and [9] for examples and [10] for a review of other multi-agent learning techniques).

Q-learning is a primitive form of learning [7] in which utility values (*Q* values) are learned for state-action pairs, absent a model of the environment. It provides a simple means for agents to learn how to act optimally in an un-
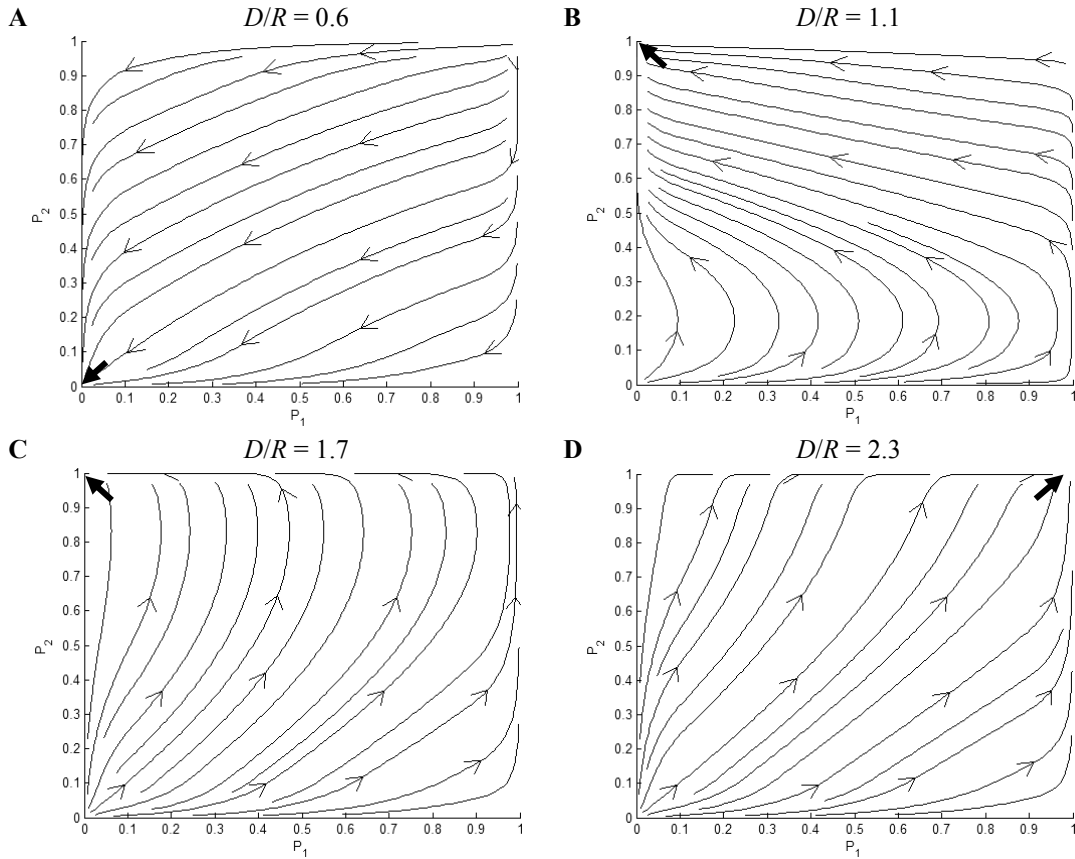

Figure 3. At any given non-equilibrium point in the strategy space, the agents have a preferred direction of moving in order to decrease their expected costs. The net direction towards which one moves in the strategy space depends on the slope of the cost functions at that point. Arrows show the net direction of movement when $n = 2$ and D/R ratio changes from 0.6 to 2.3. Each game's equilibrium is shown by a large arrow. (**A**) D/R is small. In this case, both agents ignore the intruders. (**B, C**) D/R is moderately large; the equilibrium is such that Agent 1 ignores ($p_1 = 0$) but Agent 2 elimination effort is maximum ($p_2 = 1$). (**D**) When D/R is large enough, the equilibrium changes to ($p_1 = 1$, $p_2 = 1$): the elimination effort of both agents is maximal.

known environment. At each step, a *Q*-learning agent uses its new experience to improve its long-term reward estimate by combining new information with prior experience.

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

(1) Observe the current state (in this study, there is only one).
(2) Choose and execute an action based on the *Q*-values from a set of available actions, Act_i (available actions are Act_1 = eliminate and Act_2 = ignore). The agent selects its action according to a probability given by the Boltzmann distribution:

$$p(Act_i) = \frac{e^{Q(Act_i)/T}}{\sum_{Act_k \in Actions} e^{Q(Act_k)/T}} \qquad \text{(Eq. 4)}$$

where *T*, called "temperature," adjusts the randomness of decisions.
(3) Observe the new state (for this study, this step is not necessary because there is only one state) and receive an immediate reward.
(4) Adjust *Q* value based on the action taken, *a*, using Eq. 5:

$$Q(a) \leftarrow (1-\alpha)Q(a) + \alpha(reward + \beta V)$$
$$V = \max_b Q(b) \qquad \text{(Eq.5)}$$

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

2.1. The Agent-Based Model

In mammalian livers, an absorbed toxin can be cleared by any of a sequence of hepatocytes. In order to gain insight into that process, we modeled the liver as consisting of many, parallel sets of toxin eliminating agents arranged in sequence from PV to CV in Fig. 1. Each agent used *Q*-learning algorithm to decide its clearance strategy. An agents' task was to minimize the extrahepatic damage to the organism of which they are part. We specified that agents become aware of extrahepatic tissue damage via alarm signals that are quickly released into blood by the damaged tissue. Doing so was based on the fact that hepatocytes, like immune cells, express toll-like receptors [15]. They enable cells to detect chemical alarm signals generated by damaged tissues.

3. RESULTS

Figure 4 shows the emerged strategies obtained for different values of *D/R*. When *D/R* is small, agents expend little effort to eliminate intruders. As *D/R* increases, downstream agents expend more elimination effort than upstream agents. When *D/R* is large, upstream agents begin cooperat-

ing and contribute to the elimination process, until all agents are expending maximum effort. Although the downstream agents always expend an equal or greater effort than do upstream agents, it does not mean that downstream agents actually eliminate more intruders. For example, when *D/R* = 5.0, upstream agents eliminate more intruders than do downstream agents.
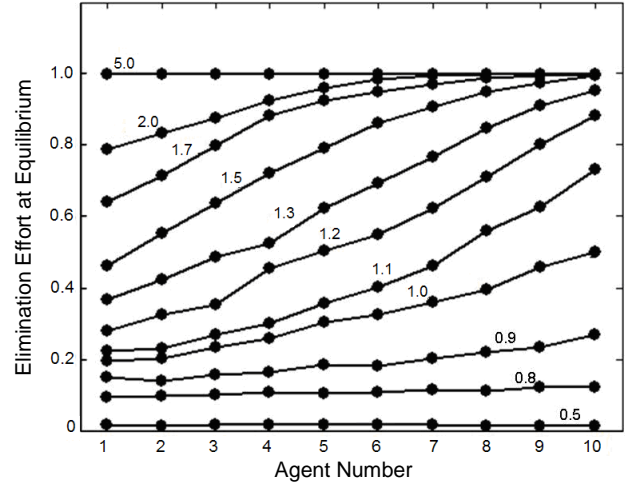

Figure 4. Equilibrium of the *n*-player game changes with *D*. Upstream is to the left (small numbers) and downstream is to the right. The average elimination strategies of 10 agents are shown for different ratios of *D/R* (shown on each curve) after 10,000 simulation steps ($maxE_i = 0.05$). When *D/R* is small, upstream and downstream agents expend little elimination effort. As *D/R* ratio increases, downstream agents expend more elimination effort than upstream agents. When *D/R* is large, upstream agents start to cooperate and contribute to the elimination process.

4. VALIDATION

Liver cells exhibit a similar behavior: they express heterogeneous, location-dependent enzyme and transporter activities to detoxify compounds. Downstream cells commit most to toxin elimination; in other words, xenobiotic metabolism is preferentially located downstream, in the perivenous region [1]. The model also suggests that the location of hepatotoxicity depends on compound toxicity: less toxic compounds are likely to damage the downstream region, whereas compounds that are more toxic are more likely to damage the upstream region. Several highly toxic compounds [11], such as TCDD (LD50 = 0.034 mg/kg), cyclochlorotine (LD50 = 2-3 mg/kg), and gossypol (LD50 = 5 mg/kg), selectively damage the upstream zone [12-14]. On the other hand, less toxic compounds such as acetaminophen (LD50 = 1295 mg/kg) selectively damage the downstream zone [11].

5. DISCUSSION

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

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

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

We speculate that hepatocytes contain the following minimum set of subsystems:
(1) One subsystem minimizes directed resource consumption by down-regulating xenobiotic-eliminating cellular components.
(2) A second subsystem receives and process alarm signals. Based on signal intensity, it upregulates xenobiotic eliminating cellular components.
(3) A third subsystem processes blood-borne signals (e.g., oxygen) to estimate hepatocyte position within a lobule. It uses this information to manage the interaction of the above two subsystems.

REFERENCES

[1] Jungermann K. 1995 Zonation of metabolism and gene expression in liver. *Histochem. Cell Biol.* **103**(2):81-91.

[2] Abu-Zahra TN, Pang KS. 2000 Effect of zonal transport and metabolism on hepatic removal: Enalapril hydrolysis in zonal, isolated rat hepatocytes in vitro and correlation with perfusion data. *Drug Metab Dispos* **28**(7):807-813.

[3] Gray MR, Tam YK. 1987 The series-compartment model for hepatic elimination. *Drug Metab Dispos* **15**(1):27-31.

[4] Lamers WH, Grange T, Moorman AF, Ruijter JM, Sassi H, Christoffels VM. 1999 A mechanistic model for the development and maintenance of portocentral gradients in gene expression in the liver. *Hepatology* **29**(4):1180-92.

[5] Daskalakis C, Papadimitriou CH. 2005 Three-player games are hard. *Electronic colloquium on computational complexity* TR05-139.

[6] Daskalakis C, Goldberg PW, Papadimitriou CH. 2006 The complexity of computing a Nash equilibrium. *Proceedings of the thirty-eighth annual ACM symposium on Theory of Computing* 71-78.

[7] Watkins, Christopher J. C. H., Dayan P. 1992 Technical note: Q-learning. *Machine learning* **8**(3):279-92.

[8] Hu J, Wellman MP. 2003 Nash q-learning for general-sum stochastic games. *Journal of machine learning research*. **4**:1039-1069.

[9] Yang E. and Gu D. 2004 Multiagent Reinforcement Learning for Multi-Robot Systems: A Survey. *Technical Report CSM-404. University of Essex*.

[10] Shoham Y, Powers R, Grenager T. 2007 If multi-agent learning is the answer, what is the question? *Artificial Intelligence* **171**(7):365-377.

[11] TOXNET US Nat'l Library of Medicine, 2008. http://toxnet.nlm.nih.gov/. Accessed 05/2008

[12] Birnbaum LS, Alcasey SK, Lucier GW, Lindros KO, Blanton J, Walker NJ, Blanton J, Lindros KO, Lucier GW, Alcasey SK and Birnbaum LS. 1999 Dose-dependent localization of TCDD in isolated centrilobular and periportal hepatocytes. *J Toxicol Sci* **52**(1):9-19.

[13] Tatsuno T, Ito E, Terao K. 1984 Liver injuries induced by cyclochlorotine isolated from penicillium islandicum. *Arch. Toxicol*. **55**(1):39-46.

[14] Lin YC, Nuber DC, Manabe S. 1991 Zone-specific hepatotoxicity of gossypol in perfused rat liver. *Toxicon*. **29**(6):787-790.

[15] Billiar TR, Vodovotz Y, Geller DA, Salyapongse AN, Liu S. 2000 Hepatocyte toll-like receptor 2 expression in vivo and in vitro: Role of cytokines in induction of rat TLR2 gene expression by lipopolysaccharide. *Shock* **14**(3):361-365.