PART I

DRUG EFFICACY AND SAFETY TECHNOLOGY
1

FOCUS ON THE FUNDAMENTALS: TOWARD BETTER THERAPEUTIC INDEX PREDICTION

JINGHAI J. XU AND LI J. YU

Contents

1.1 Introduction 3
1.2 General Approaches in Predicting the Therapeutic Index 5
1.3 Predicting Efficacious Exposure 6
  1.3.1 Clinically Relevant Endpoint 7
  1.3.2 Simulated In Vivo Milieu 9
  1.3.3 Kinetics Rule 12
1.4 Predicting Side Effects/Toxicity Exposure 13
  1.4.1 Cardiac Toxicity 13
  1.4.2 Hepatotoxicity 14
  1.4.3 Muscle Toxicity 18
  1.4.4 Is Dose Escalation a Common Theme in Estimating Toxic Exposure? 19
1.5 Future Perspectives 20
References 21
Acknowledgment 28

1.1 INTRODUCTION

“What is there that is not poison? Everything can be poison. What differentiates a poison from a medicine is its dose” [1]. This is the famous revelation of Paracelsus (1493–1541), the father of modern medicine. Modern
pharmaceutical scientists have refined the concept of dose with concentration, and added a time variable [2]. Successful drug therapy relies on not drug potency per se but on a wide enough safety margin between toxicity and efficacy, the so-called sufficient therapeutic index (TI) [3]. This safety margin is exemplified by the lipid-lowering drug class hydroxymethylglutaryl-CoA (HMG-CoA) reductase inhibitors known as statins. Cerivastatin is a potent statin with therapeutic exposure levels that are an order of magnitude lower than other statins. However, it was withdrawn from the market after widespread prescription because of a significantly higher incidence and severity of muscle toxicity [4]. The challenge facing all of us in the pharmaceutical research and development field is how to establish robust screening paradigms and predictive strategies to (1) proactively identify drugs with narrow therapeutic indexes, before they are administered to humans, and (2) select drug candidates with sufficient safety margins that can withstand the test of time and widespread use.

Traditionally drug safety evaluation is conducted in the later stages of development when the entire development program is focusing on a single new chemical entity (NCE). In this mode of compound testing, the safety evaluation of the NCE is composed of preclinical animal toxicology, pathology, and safety pharmacology studies followed by phased human clinical tests (i.e., phases I, II, III, and postmarketing surveillance). In the event that severe toxicity occurs without a sufficient TI, it can lead to abandonment of such development programs or severe restrictions on the utility of such an NCE (Figure 1.1). It is now widely recognized that drug toxicity is the leading cause of drug candidate attrition in development [5], as well as drug withdrawals

![Traditional Discovery and Development](image)

*Q: Why not screen against the end goal: therapeutic index (TI)?*

**Figure 1.1** Role of therapeutic index (TI) screening in drug discovery and development.
from the market after they were initially approved by the drug regulatory agencies [6]. Approximately one-third of drug attrition is due to preclinical drug toxicity or clinical safety [7]. These statistics, however, may underestimate the contribution of toxicity to overall drug attrition. “Lack of efficacy” has been the other leading cause of drug attrition. However, among the drug candidates that have exhibited poor efficacy, there is evidence that for a significant number of these candidates, the clinical dosing regimen was limited or restricted by dose-limiting toxicity. In other words, the proof of clinical efficacy could not be fully achieved if an NCE did not have a sufficient TI (e.g., [8]). Hence an additional fraction of the “lack of efficacy” group also contained an underlying drug safety cause. If one would introduce a category called “lack of sufficient therapeutic index,” this category would certainly occupy the largest cohort of drug attrition [9].

If lack of sufficient TI is the major hurdle of pharmaceutical research and development (R&D), the obvious question is, why not focus on this key parameter by applying emerging technologies to screen for improved TI in key transitioning steps during prolonged R&D programs? If this screening strategy is applied early enough in the drug discovery stage (e.g., right after cell-based efficacy screens), the data could help the drug discovery teams proactively identify and select drug candidates with an increasing probability of improved TI in vivo. The remainder of this chapter will discuss emerging technologies that could make this happen, and provide the authors’ perspectives on challenges in the field.

1.2 GENERAL APPROACHES IN PREDICTING THE THERAPEUTIC INDEX

Predicting a therapeutic index is intrinsically a systems pharmacology and systems toxicology challenge. This exercise no longer deals with a single enzyme or receptor, but with a multiple-component system that includes competing endogenous ligands, co-factors, off-targets, multiple cell types, organs, and tissues. Many scientists view systems biology as an “integrated” omics. While there is a need to integrate all the omic data to generate a “pathways view” of biology, this per se does not allow quantitative prediction of a therapeutic index. To translate biological knowledge into an effective therapy, one needs to quantitatively predict the fundamental attributes of drugs, for example, the intrinsic TI of a drug and different ways to optimize TI in vivo. This necessitates an accurate prediction of both efficacious and toxic concentrations (Figure 1.2). Hence the in vitro–in vivo correlations for cellular models, biomarkers, and concentrations need to be established for both drug efficacy and drug toxicity. Furthermore drug disposition and pharmacokinetic (PK) properties need to be taken into consideration for the final calculation of therapeutic index. In this sense, systems pharmacology and systems toxicology build on basic biological knowledge derived from systems biology, but
FOCUS ON THE FUNDAMENTALS

1.3 PREDICTING EFFICACIOUS EXPOSURE

A major challenge in predicting drug efficacious exposure is the apparent “loss of efficacy” through the translation of model systems. Such “loss of efficacy” may happen when transitioning from the isolated enzyme or receptor screens to whole cell assays, from whole cell assays to animal models, from animal models to human proof-of-concept (POC) studies, from human POC studies to large-scale and/or longer term human patient trials. Since such “loss of efficacy” leads to negative results, and negative results are typically considered less interesting to publish, our systematic understanding of the mechanisms behind such a “loss of efficacy” are relatively poor. This is unfortunate since more productive strategies of drug R&D cannot be devised unless we integration that with in vitro–in vivo correlation, PK measurements, pharmacodynamic prediction, and knowledge about mechanisms of chemical toxicity and its reversibility.

**Figure 1.2** Therapeutic index (TI) screening requires the quantitative prediction of clinical efficacy and safety by a combination of cellular models, biomarkers, and in vivo relevant concentrations.
understand such mechanisms. After all, the number of negative results by far exceeds the successful NCE launches in the R&D endeavor [7].

While there are few systematic studies on the mechanisms of such apparent “loss of efficacy,” anecdotally several reasons have been postulated: (1) preclinical efficacy models do not correlate with clinical diseases, (2) preclinical efficacy biomarkers do not correlate with disease regression, and (3) in vitro effective concentrations do not correlate with in vivo concentrations. Fundamentally too much emphasis on the isolated disease target and the reductionist approach of drug discovery has driven us away from a more holistic understanding of disease pathophysiology in an “open system” in vivo. Translational medicine, which focuses on the quantitative correlation between in vitro and in vivo, between animals and humans, aims to bridge this gap. In this section we will use case studies to highlight certain key principles of such translation and present emerging technological approaches that could address this challenge.

1.3.1 Clinically Relevant Endpoint

An example of a “loss of efficacy” during preclinical to clinical transition is the case of acyl-CoA cholesterol acyltransferase (ACAT) inhibitors. ACAT converts free cholesterol to cholesterol esters; hence it catalyses the reverse reaction of cholesteryl ester hydrolase (CEH). The original hypothesis was that if ACAT in peripheral cells such as foamy macrophages was inhibited, more cholesterol would stay in the form of free cholesterol instead of cholesterol esters. Since it was thought that only free cholesterol could leave the foamy macrophages, it was hypothesized that inhibition of ACAT could drive the efflux of cholesterol from these cells, and ultimately up-regulate the rate of reverse cholesterol transport (RCT) back to liver for subsequent clearance (Figure 1.3). In cellular efficacy models it was found that ACAT inhibitor

![Figure 1.3](image_url)
significantly inhibited cholesterol esterification in rat aortic smooth muscle cells and macrophages. It reduced cholesterol ester by 97% and increased the cellular free cholesterol by approximately twofold [10]. In the cholesterol-fed rabbit model, the ACAT inhibitor was shown to prevent the formation of atherosclerosis and even accelerate its regression [11]. In fact several ACAT inhibitors that were tested in this animal model have showed some degree of anti-atheroscerotic effects [11–21].

To date two ACAT inhibitors, avasimibe and pactimibe, have been tested in large-scale human clinical trials. However, both trials showed lack of efficacy or even exacerbated the atheroma outcome compared to the current standard of care (i.e., statin therapy). In the Avasimibe and Progression of Lesions on Ultrasound (A-PLUS) trial, over 600 patients with coronary atherosclerosis were randomized to receive a statin or statin plus avasimibe for a period of two years [22]. The primary outcome was plaque burden, as measured by intravascular ultrasonography (IVUS) in all patients. At the end of the trial, the percent atheroma volume increased or worsened by 0.4% to 1% compared with placebo (statin alone) in a dose-dependant manner in the dose-escalating avasimibe plus statin groups [22]. In the ACAT Intravascular Atherosclerosis Treatment Evaluation (ACTIVATE) trial, over 400 patients with coronary atherosclerosis were randomized to receive statin or statin plus pactimibe, another ACAT inhibitor, for 18 months [23]. The primary outcome was also plaque burden, as measured by IVUS. In this study the atheroma volume in the most diseased 10-mm segment regressed by 3.2 mm$^3$ in the placebo (i.e., statin) group, as compared with a decrease of only 1.3 mm$^3$ in the pactimibe group ($P = 0.01$) [23].

The fact that both trials with two different ACAT inhibitors showed less than the expected atheroma outcomes compared to statin alone, showed a lack of correlation between the clinically relevant endpoint, in this case IVUS imaging, and previously measured endpoints in vitro (e.g., inhibition of ACAT per se). It is possible that direct measurements of the rate of cholesterol efflux from cholesterol-filled human macrophages could be a much more clinically relevant endpoint to measure in vitro. Indeed pre-incubation of THP-1 macrophages with atorvastatin, a widely-prescribed statin, dose dependently stimulated cholesterol efflux to apolipoprotein AI (apoAI) and high density lipoprotein (HDL) [24]. This was confirmed by ex vivo cellular cholesterol efflux studies in patients treated with atorvastatin [25]. In the future such cholesterol efflux studies (both in vitro and ex vivo) should be conducted to compare the effects of the statin alone with the statin plus an ACAT inhibitor. Since statin treatment already inhibits cholesteryl ester accumulation in macrophages challenged with atherogenic hypertriglyceridemic very low density lipoproteins (VLDL) [24], it is possible that any further reduction of cholesteryl ester content by avasimibe or pactimibe did not result in any additional increase in the rate of reverse cholesterol transfer (i.e., ACAT no longer being the rate-limiting step of reverse cholesterol transfer).
An analogous situation exists in the debate about which is more important in vivo, high-density lipoprotein (HDL) concentration or HDL function? Since current clinically accepted measurements are IVUS imaging of plaque burden in phase 2 or 3 trials, the HDL function on reverse cholesterol transfer is likely to be a more important translational biomarker. Indeed the ApoA-1 (Milano) carriers originally identified in northern Italy have very low HDL levels [26]. But they also have exceptional longevity with apparent protection from atherosclerosis [26,27]. The putative “HDL functions” can be measured both in vitro and ex vivo by (1) the ability to induce cholesterol efflux in aortic smooth muscle cells and macrophages, (2) the apoA-1 level, (3) paraoxonase 1 (PON1) activity, (4) reactive oxygen species (ROS) content, and (5) the ability to prevent low-density lipoprotein (LDL) oxidation. Since many clinically effective anti-atherosclerosis medicines have been shown to improve one or more of these “HDL functions,” it is anticipated that a composite index of “HDL function” is a better biomarker for atherosclerosis risk than HDL levels [28–30]. It is of interest to note that physiologically relevant assays to measure all of these five HDL functions are available [30]. Therefore future HDL-targeted therapies should include these functional assays as part of their screening paradigm.

1.3.2 Simulated In Vivo Milieu

An apparent “loss of efficacy” can occasionally be obscured by a lack of in vitro–in vivo concentration correlation. In the current drug discovery paradigm, initial lead compounds are often identified from in vitro high-throughput screens (HTS), for example, an inhibition assay against a target enzyme or receptor. It is often hoped that the in vitro potency (e.g., IC\textsubscript{50}) measured can be used to predict the in vivo pharmacological activity (e.g., EC\textsubscript{50}). When such a correlation does not exist, there is added uncertainty as to the validity of the pharmacological target and/or the in vitro screening strategy. Also the HTS cannot be relied upon to drive the structure activity relationship (SAR), and quantitative predictions of TI become impossible to derive. Hence establishment of a correlation between in vitro potency and in vivo activity is crucial in selecting drug candidates with an optimal therapeutic window.

Theoretically the average free efficacious concentration at the steady state in vivo should be correlated with the free efficacious concentration determined from an in vitro assay as described by Recant and Riggs [31]. The Recant equation constitutes a fundamental free ligand hypothesis and has been supported by many researchers [32–35]. In practice, however, this relationship is often obscured or confounded due to a variety of factors. For example, nonphysiological conditions and nonspecific binding within in vitro systems may yield an inaccurate estimate of true intrinsic potency. In addition complex PK–PD relationships arising from indirect effects or target site disequilibrium may result in inappropriate determination of potency in vivo.
This was the case for the initial lack of correlation between in vitro potency and in vivo activity for human liver glycogen phosphorylase a (GPa) inhibitors [36]. Inhibition of liver GPa blocks the glycogenolysis pathway and therefore leads to reduction of hepatic glucose production [37,38], an effect that may be beneficial for treating type 2 diabetes mellitus. Initially, among 13 GPa reversible inhibitors (GPIs), no obvious correlation was observed between in vitro GPa enzyme IC\textsubscript{50}s and any in vivo measurements such as the minimum efficacious dose (MED), or plasma or liver total and free drug concentrations at MED. It was hypothesized that the IC\textsubscript{50} values obtained from HTS was probably determined under nonphysiological conditions because of the absence of adequate amounts of important cofactors or modulators present in vivo, such as AMP, ATP, and other unknown factors. Indeed there is evidence that some of these cofactors affect the binding affinity of a GPI to a GPa in vivo [39–41].

Theoretically the in vivo activity should be driven by the amount of enzyme bound to the administered inhibitor, which in turn is determined by the intrinsic dissociation constant (\(K_d\)) and the free enzyme concentration at the target site. However, accurate measurement of \(K_d\) for a target enzyme under physiological conditions can be complicated and time-consuming [35,42]. A robust method was developed to determine the intrinsic \(K_d\) of inhibitors for a purified human enzyme in the presence of organ homogenate using a previously validated 96-well equilibrium dialysis apparatus [43,44]. These conditions were supposed to mimic those in vivo better than the conventional assay [38] because most, if not all, cofactors are present in the organ homogenate (in this case, liver). The method utilizes three 96-well equilibrium dialysis apparatus in the following configurations (Figure 1.4): (1) liver homogenate against the target enzyme (in this case, GPa), (2) liver homogenate against the buffer, and (3) the target enzyme against the buffer. In the parallel dialysis setups 1 and 2, small cofactors that exist in the liver homogenate can pass the dialysis membrane and reach the enzyme chamber or buffer chamber, respectively. To facilitate rapid equilibrium, the inhibitor was added to both chambers at concentration of 250 ng/mL (\(\sim\)0.5 \(\mu\)M). At the end of dialysis (7 h), total concentrations of inhibitor at both sides of the chambers were determined by LC/MS/MS measurements [36].

By this highly parallel equilibrium dialysis technology, a significantly improved in vitro–in vivo correlation was found between the free \(K_d\) determined in the presence of liver homogenate and the free inhibitor liver concentration at the MED [36]. In addition, by this method, the concentration ratio of an inhibitor bound to the target enzyme over the total inhibitor concentration in the liver \((\text{[I]}_{\text{b-GPa}}/\text{[I]}_{\text{t-liver}}\) was estimated. This ratio has several important implications: (1) it can predict the total liver exposure in vivo at the MED, (2) it is useful in estimating the amount of target enzyme bound to an inhibitor at the MED, (3) the higher this ratio is, the more “selective” an inhibitor is for the target enzyme as opposed to the other “off-target” components in liver, and (4) this ratio represents the target efficiency of local drug
disposition in this particular organ. One may extend this methodology to include a second organ, in order to estimate the relative drug disposition between two different organs. Such local drug disposition studies may have important implications in predicting TI. After all, it is the local drug concentration and ultimately local enzyme/receptor occupancy that drives drug efficacy.

This example highlights the importance of evaluating in vitro drug potency in the relevant in vivo milieu. This can be done using the appropriate endogenous ligands and co-factors at in vivo relevant concentrations. When such information is not known, whole organ lysates may be used instead. In general, a better understanding of local drug disposition studies within the target organ is an important consideration for drugs that target intracellular enzymes.
1.3.3 Kinetics Rule

The previous example relies on the thermodynamic dissociation constant, $K_d$ (or $K_i$), as the primary measure of drug–target potency. However, there is also an underlying need to study kinetic properties of drug candidates using in vitro approaches, and the resulting data to make efficacy predictions [45–47]. In particular, since every drug will eventually be cleared from the body, the dissociation kinetics of a drug candidate from its target enzyme needs to be evaluated and placed into the context of its pharmacokinetic and disposition properties. Until now, few researchers have made the effort to systematically study compound SAR on the basis of on-rate and off-rate determinations, and fewer still have translated this information into in vivo effects [48].

One exception is an in vitro–in vivo correlation study on three neurokinin 1 receptor (NK-1R) antagonists [49]. In preclinical animal models the NK-1R was shown to be involved in emesis, asthma, psychiatric disorders, gastrointestinal disorders, pain, migraine, inflammation, and urinary bladder disorders (reviewed by [50]). However, to date only aprepitant has reached the market for treatment of chemotherapy-induced emesis, despite prolonged and extensive efforts by many pharmaceutical companies [51]. In a proof of concept study three different NK-1R antagonists were compared with respect to their functional dissociation kinetics to NK-1R. The study was performed in U373MG human astrocytoma cells endogenously expressing the human NK-1R. Substance P-induced mobilization of intracellular calcium ($\text{Ca}^{2+}$) was measured by the fluorescent calcium-sensitive dye Fluo-4, which provided the kinetic readouts of the functional interaction to NK-1R. Cells were pre-loaded with 4μM Fluo-4 together with 10nM of each antagonist or buffer. After 30 minutes of incubation, the cells were washed three times in assay buffer, and the “functional dissociation kinetics” of each antagonist from the NK-1R was determined by the residual Fluo-4 dye fluorescence upon substance P stimulation at various time points post-drug removal [49]. The inhibitory effect of CP-99994 was abolished within 30 minutes, whereas for ZD6021, 50% inhibition still persisted after 60 minutes. In contrast, aprepitant produced maximal inhibition lasting more than an hour. The “functional dissociation kinetics” measurement correlated very well with the duration of efficacy in the preclinical in vivo animal models of NK-1R antagonism. Slow functional reversibility of aprepitant was associated with long-lasting efficacy in vivo, whereas the efficacy of compounds with rapid reversibility closely mimicked their pharmacokinetic profiles in vivo [49].

The three examples highlighted here serve to remind us that drugs in vivo interact with complex systems including endogenous ligands and co-factors, other “off-targets,” and other binding parties. A successful prediction of in vivo efficacious concentration requires a renewed thinking of in vivo-relevant endpoints, cell milieu, concentrations, dissociation kinetics, pharmacokinetic and disposition properties. Key considerations to correlate in vitro potency to in vivo activity include (1) the in vitro endpoint measured is not limited to the
known activity of the target enzyme or receptor but resembles relevant in vivo clinical endpoints, (2) the in vitro milieu utilized closely resembles the in vivo milieu, (3) the in vitro dissociation kinetics is studied and integrated with the in vivo pharmacokinetic measurements and simulations. The approaches presented here can be applied to many drugs in pharmacology studies where there is a need to use preclinical efficacious concentrations ($C_{\text{eff}}$) to predict in vivo $C_{\text{eff}}$ and duration of effects.

Finally, it is important to reiterate that the drug efficacy consideration alone is not sufficient for a reliable TI prediction. For example, if a drug-induced side effect is directly linked to hyperantagonism of a target enzyme (as is the case for cerivastatin), a slow dissociation rate and long residence time can potentially exacerbate such a side effect (especially upon repeated drug dosing). While it is likely that poor efficacy could stem from rapid drug dissociation from its efficacy target, toxicity could result from slow drug dissociation from its “toxicity target.” This can explain why drug-induced organ toxicity typically takes repeated dosing and manifests over longer period of time, as intracellular drug concentrations can reach a toxic level. Therefore it is important to assess whether toxicity is likely to occur at supratherapeutic levels during drug candidate selection.

1.4 PREDICTING SIDE EFFECTS/TOXICITY EXPOSURE

We now turn to the prediction of drug exposure levels where toxicity is likely to be experienced in vivo, as therapeutic index is defined as the ratio between toxic and efficacious concentrations. Major dose-limiting toxicities encountered by pharmaceutical products include cardiac toxicity [52], hepatotoxicity [53,54], muscle toxicity [55], hematotoxicity [56], genetic toxicity [57], and teratogenecity [58]. While many studies have been published on a particular drug and its associated organ toxicity, what is still needed at the drug discovery stage is a robust quantitative model that can predict across multiple drugs the exposure of a toxic drug level compared to a nontoxic drug level. Due to space limitations not all organ toxicities will be covered in detail in this chapter. The examples provided in this chapter will instead aim to highlight the key strategies and challenges in predicting toxicity exposure.

1.4.1 Cardiac Toxicity

Human life heavily depends on a functional human heart. Hence it is critical to predict and evaluate the safety margin for cardiac side effects for all drugs. In particular, the potential proarrhythmic risk of new drug candidates is a major subject of concern and needs to be carefully addressed before treatment of human volunteers or patients takes place [59]. The prolongation of the time interval between Q and T waves in an electrocardiogram (i.e., QT prolongation), is now a well-recognized biomarker for a life-threatening form of
arrhythmia, called torsade de pointes (or twisting of points in an electrocardiogram) [60]. The molecular mechanisms of drug-induced QT prolongation have been well characterized. Many drugs that prolong the QT interval also block the hERG (human ether-a-go-go-related gene) potassium channel (I_{Kr} channel) (reviewed by [61,62]). Drug regulators now routinely require ligand binding studies for the I_{Kr} channel, and thorough QT assessment in preclinical species and in clinical trials [63]. Indeed the I_{Kr} channel binding assay and QT prolongation evaluations are probably the best characterized drug safety tests developed in the past decade. The collaborative work among scientists from several pharmaceutical companies suggested a provisional safety margin of 30 times the drug’s efficacious exposure for in vitro hERG I_{Kr} channel binding assays [64]. However, significant gaps and challenges still exist. They include (1) estimation of efficacious exposure (related to the “loss of efficacy” when transitioning from one drug efficacy model to the next) and (2) whole systems’ understanding of the interactions among multiple ion channels of the heart.

In one example, a compound that was inactive at hERG channels produced no significant QT changes in preclinical dog studies or in phase I trials with healthy human volunteers. However, prolongation of QT occurred in osteoarthritis and diabetic neuropathy patients in phase II trials, as a result of a significantly higher systemic exposure in these patients, compared with healthy volunteers [65]. This case study illustrated the importance of verifying exposure values and hence safety margins as new clinical pharmacokinetic data are available.

Small molecules can interact with not just one ion channel, but multiple ion channels of the heart. Such interactions can either “cancel” or exacerbate its proarrhythmic risk. More thorough studies should be performed to systematically characterize the interactions of such small drug molecules with several well-known cardiac ion channels. The systems biology modeling efforts to integrate the sustained I_{(Na+)} and L-type I_{(Ca2+)}, in addition to I_{Kr} inhibition values hold promise to provide more holistic predictions of cardiac tissue’s arrhythmic outcome in the future [66,67].

1.4.2 Hepatotoxicity

Since the liver is a highly perfused and the “first-pass” organ for any orally administered xenobiotic, it is a frequent site of toxicity of pharmaceuticals in humans [53,68]. Indeed drug-induced liver injury (DILI) is the number one reason why drugs were not approved in the first place, and why some of them were withdrawn from the market after approval [9]. The physiological location and drug-clearance function of the liver dictate that for an orally administered drug, the drug exposure or drug load that the liver “sees” is higher than what is being measured systemically at peripheral blood [69]. Hence we hypothesized that for DILI, a higher safety margin than QT prolongation may be needed.
To test this hypothesis, we assembled a list of more than 300 drugs and chemicals with a classification scheme based on clinical data for hepatotoxicity. Our DILI positive drugs include those (1) withdrawn from the market mainly due to hepatotoxicity (e.g., troglitazone [70]), (2) not marketed in the United States due to hepatotoxicity (e.g., nimesulide [71]), (3) receiving black box warnings from the FDA due to hepatotoxicity (e.g., dantrolene [72]), (4) marketed with hepatotoxicity warnings in their labels (e.g., zileuton [73]), (5) others (mostly old drugs) that have well-known associations with liver injury and have a significant number (>10) of independent clinical reports of hepatotoxicity (e.g., diclofenac [74]). Drugs that do not meet any of the positive criteria above are classified as DILI negatives. Since every drug can exhibit some toxicity at high enough exposure (i.e., the notion of “dose makes a poison” by Paracelsus), we searched therapeutic exposure levels through a variety of databases (Physicians’ Desk Reference, PubMed, Pharmapendium™, Prous™) and collated the therapeutically active average plasma maximum concentration ($C_{\text{max}}$) values upon single-dose administration at commonly recommended median therapeutic doses.

To evaluate these drugs for their potential to induce hepatocyte damage, we utilized primary human hepatocytes cultured in a sandwiched configuration and multi-parameter image-based technology. Human hepatocytes cultured in the sandwiched configuration express liver-specific metabolizing enzymes [75–77], uptake and efflux transporters [78–80], and predict drug clearance via the hepatobiliary route [81]. Cryopreserved human hepatocytes were obtained commercially from CellzDirect (http://www.cellzdirect.com/). The cells were plated on collagen-coated 96-well plates (BD Biosciences) in hepatocyte plating medium (Dulbecco’s Minimal Essential Medium with 5% fetal bovine serum; all media obtained from CellzDirect). Upon cell attachment, the medium was changed to hepatocyte culturing medium (Williams E medium). On the second day, the hepatocytes were sandwiched by applying an overlay of Matrigel™ (BD Biosciences). On the third day, the cells underwent a medium change with hepatocyte culturing medium. On the fourth day, the cells were treated overnight with the compound of interest or vehicle (0.1% DMSO). All compounds were initially solubilized in DMSO and diluted in culturing medium containing 5% fetal bovine serum to a final DMSO concentration of 0.1%. After 24 hours of incubation (37°C, 5% CO₂, 100% humidity) media were removed and the cells were stained by fluorescent probes in the same culturing medium lacking serum. The fluorescent probes were tetramethyl rhodamine methyl ester for mitochondrial membrane potential (TMRM; 0.02μM, 1 h), 1,5-bis[2-(dimethylamino)ethyl]amino]-4,8-dihydroxyanthracene-9,10-dione for nuclei and lipids DNA (DRAQ5; 45μM, 30 min), 5-(and-6)-chloromethyl-2′,7′-dichlorodihydrofluorescein diacetate acetyl ester for reactive oxygen species (CM-H₂DCFDA; 10μM, 30 min), and finally monochlorobimane for glutathione (mBCl; 80μM, 5 min). Automated live-cell multispectral image acquisition was performed on a Kinetic Scan Reader (Cellomics (http://www.cellomics.com/) using a 20× objective.
and an XF93 filter. The fluorescent images were captured according to the excitation and emission wavelengths of each probe. To capture enough cells (>500) for analysis, six image fields starting from the center of each well were collected. Image analysis was performed using ImagePro Plus (Media Cybernetics, Bethesda, MD).

It was found that the 100-fold \( C_{\text{max}} \) scaling factor represented a reasonable threshold to differentiate safe from toxic drugs, for an orally dosed drug and with regard to hepatotoxicity (Table 1.1). The first 11 drugs in Table 1.1 were known to cause idiosyncratic hepatotoxicity in humans. These drugs induced changes in the human hepatocyte imaging assays in concordance with many of their known mechanisms of hepatotoxicity. Perhexiline was reported to induce nonalcoholic steatohepatitis (NASH) in humans [82]. The mechanism of perhexiline-induced NASH involves mitochondrial injury, which causes steatosis due to impaired beta-oxidation of fatty acids, and leads to the generation of reactive oxygen species and ATP depletion [82,83]. In our imaging assay, perhexiline increased ROS and lipid intensity, and decreased mitochondrial membrane potential, GSH content, and GSH area. Troglitazone, a diabetic drug that was withdrawn from the market due to idiosyncratic liver injury, was known to cause mitochondrial damage in the literature [84,85], and completely depleted mitochondrial membrane potential in our assay (Table 1.1). In addition troglitazone also increased lipid intensity, and decreased GSH content in the hepatocytes. Nefazodone, an antidepressant that was withdrawn from the market due to hepatotoxicity [86], exhibited substantially higher frequency and severity of hepatic injury compared to other prescribed antidepressants. The hepatocyte imaging assay points to mitochondrial damage as a potential target underlying its side effects. Tetracycline, an antibiotic that was frequently associated with liver injury, was reported to induce oxidative stress both in vitro and in vivo [87-89]. As expected, the ROS signal was substantially increased in our assay. Tetracycline is also known to cause steatosis and affect mitochondria both in vitro and in vivo [90,91], and this was confirmed in the imaging assay by measurements of lipid and mitochondrial membrane potential. Nimesulide is a nonsteroidal antiinflammatory drug (NSAID) associated with a higher risk of hepatotoxicity compared to other NSAIDs [92]. Even though in vitro studies have suggested mitochondria as a potential target of toxicity, this has not been substantiated by in vivo studies [93]. Indeed, in human hepatocytes, nimesulide increased ROS and lipid intensity without affecting mitochondrial health. Likewise, sulindac and diclofenac, the other two NSAIDs with a higher risk of hepatotoxicity in the NSAID class of drugs [92], also increased ROS in our model, suggesting oxidative stress as a common mechanism of NSAID-induced liver injury (Table 1.1). Our assay technology also identified mechanisms of toxicity for those drugs that caused severe clinical DILI, but with unknown mechanisms thus far. Zileuton, a 5-lipoxygenase inhibitor approved for the treatment of asthma in adults and children, is known to cause idiosyncratic hepatocellular damage [73]. The imaging assay results suggested an increased level of oxidative stress and intracellular lipids.
TABLE 1.1  High Content Image Analysis Output from the Human Hepatocyte Imaging Assay Technology (HIAT)

<table>
<thead>
<tr>
<th>Drug</th>
<th>DILI Label</th>
<th>Nuclei Count (&lt;0.4 = Positive)</th>
<th>Nuclei Area (&lt;0.4 = Positive)</th>
<th>ROS Intensity (&gt;2.5 = Positive)</th>
<th>TMRM Intensity (&lt;0.4 = Positive)</th>
<th>Lipid Intensity (&gt;2 = Positive)</th>
<th>GSH Content (&lt;0.4 = Positive)</th>
<th>GSH Area (&lt;0.4 = Positive)</th>
<th>Test Score (Logical OR of All Results)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perhexiline</td>
<td>P</td>
<td>0.96</td>
<td>1.07</td>
<td>24.0</td>
<td>0.04</td>
<td>2.48</td>
<td>0.01</td>
<td>0.02</td>
<td>P</td>
</tr>
<tr>
<td>Troglitazone</td>
<td>P</td>
<td>0.71</td>
<td>0.85</td>
<td>0.00</td>
<td>0.01</td>
<td>2.90</td>
<td>0.01</td>
<td>0.02</td>
<td>P</td>
</tr>
<tr>
<td>Nefazodone</td>
<td>P</td>
<td>0.45</td>
<td>0.52</td>
<td>0.00</td>
<td>0.04</td>
<td>1.81</td>
<td>0.01</td>
<td>0.03</td>
<td>P</td>
</tr>
<tr>
<td>Tetracycline</td>
<td>P</td>
<td>1.04</td>
<td>1.06</td>
<td>430</td>
<td>0.10</td>
<td>2.90</td>
<td>1.13</td>
<td>0.82</td>
<td>P</td>
</tr>
<tr>
<td>Nimesulide</td>
<td>P</td>
<td>0.96</td>
<td>0.92</td>
<td>27.3</td>
<td>1.43</td>
<td>2.86</td>
<td>0.88</td>
<td>0.98</td>
<td>P</td>
</tr>
<tr>
<td>Sulindac</td>
<td>P</td>
<td>1.14</td>
<td>1.08</td>
<td>14.2</td>
<td>0.69</td>
<td>0.79</td>
<td>0.63</td>
<td>0.81</td>
<td>P</td>
</tr>
<tr>
<td>Zileuton</td>
<td>P</td>
<td>1.08</td>
<td>1.01</td>
<td>4.83</td>
<td>1.32</td>
<td>3.46</td>
<td>0.92</td>
<td>0.67</td>
<td>P</td>
</tr>
<tr>
<td>Labetalol</td>
<td>P</td>
<td>0.89</td>
<td>0.81</td>
<td>4.88</td>
<td>0.70</td>
<td>5.10</td>
<td>1.41</td>
<td>0.97</td>
<td>P</td>
</tr>
<tr>
<td>Diclofenac</td>
<td>P</td>
<td>0.61</td>
<td>0.70</td>
<td>35.3</td>
<td>0.98</td>
<td>1.86</td>
<td>1.91</td>
<td>1.22</td>
<td>P</td>
</tr>
<tr>
<td>Chlorzoxazone</td>
<td>P</td>
<td>0.48</td>
<td>0.74</td>
<td>8.39</td>
<td>0.78</td>
<td>1.70</td>
<td>1.01</td>
<td>1.65</td>
<td>P</td>
</tr>
<tr>
<td>Dantrolene</td>
<td>P</td>
<td>0.72</td>
<td>0.78</td>
<td>8.54</td>
<td>1.23</td>
<td>1.58</td>
<td>1.57</td>
<td>1.17</td>
<td>P</td>
</tr>
<tr>
<td>Amitriptyline</td>
<td>N</td>
<td>0.96</td>
<td>1.00</td>
<td>0.30</td>
<td>1.26</td>
<td>1.36</td>
<td>1.06</td>
<td>0.97</td>
<td>N</td>
</tr>
<tr>
<td>Pioglitazone</td>
<td>N</td>
<td>1.13</td>
<td>1.11</td>
<td>0.88</td>
<td>1.69</td>
<td>1.38</td>
<td>0.57</td>
<td>0.87</td>
<td>N</td>
</tr>
<tr>
<td>Rosiglitazone</td>
<td>N</td>
<td>0.66</td>
<td>0.79</td>
<td>0.44</td>
<td>1.04</td>
<td>1.66</td>
<td>1.52</td>
<td>1.38</td>
<td>N</td>
</tr>
<tr>
<td>Primidone</td>
<td>N</td>
<td>0.96</td>
<td>0.91</td>
<td>0.62</td>
<td>0.90</td>
<td>1.19</td>
<td>0.75</td>
<td>0.95</td>
<td>N</td>
</tr>
<tr>
<td>Aspirin</td>
<td>N</td>
<td>1.04</td>
<td>1.00</td>
<td>0.58</td>
<td>1.25</td>
<td>0.87</td>
<td>0.62</td>
<td>0.94</td>
<td>N</td>
</tr>
<tr>
<td>Penicillin</td>
<td>N</td>
<td>0.91</td>
<td>0.98</td>
<td>0.29</td>
<td>1.31</td>
<td>1.32</td>
<td>0.67</td>
<td>1.07</td>
<td>N</td>
</tr>
<tr>
<td>Melatonin</td>
<td>N</td>
<td>0.97</td>
<td>0.96</td>
<td>0.68</td>
<td>0.95</td>
<td>0.93</td>
<td>0.91</td>
<td>1.00</td>
<td>N</td>
</tr>
<tr>
<td>Nadolol</td>
<td>N</td>
<td>0.95</td>
<td>0.99</td>
<td>0.44</td>
<td>2.07</td>
<td>0.92</td>
<td>0.91</td>
<td>1.04</td>
<td>N</td>
</tr>
<tr>
<td>Ketotifen</td>
<td>N</td>
<td>1.02</td>
<td>1.01</td>
<td>0.60</td>
<td>1.00</td>
<td>0.88</td>
<td>0.90</td>
<td>0.94</td>
<td>N</td>
</tr>
<tr>
<td>Paromomycin</td>
<td>N</td>
<td>1.00</td>
<td>1.02</td>
<td>1.02</td>
<td>1.66</td>
<td>0.93</td>
<td>1.23</td>
<td>0.99</td>
<td>N</td>
</tr>
<tr>
<td>Sumatriptan</td>
<td>N</td>
<td>1.13</td>
<td>1.10</td>
<td>0.26</td>
<td>0.64</td>
<td>0.86</td>
<td>0.78</td>
<td>0.87</td>
<td>N</td>
</tr>
<tr>
<td>Famotidine</td>
<td>N</td>
<td>1.20</td>
<td>1.15</td>
<td>0.47</td>
<td>1.15</td>
<td>0.99</td>
<td>0.96</td>
<td>0.84</td>
<td>N</td>
</tr>
<tr>
<td>Fluoxetine</td>
<td>N</td>
<td>1.05</td>
<td>1.07</td>
<td>1.46</td>
<td>0.79</td>
<td>0.84</td>
<td>0.59</td>
<td>0.91</td>
<td>N</td>
</tr>
</tbody>
</table>

Note: In the “DILI Label” column, P = DILI positive, or hepatotoxic in humans; N = DILI negative, or nonhepatotoxic in humans according to the drug classification scheme described in the main text (see page 15). In the HIAT results columns, the test “positives” were produced by the following thresholds: reduction in nuclei count, nuclei area, TMRM intensity, GSH content, and GSH area to 40% or less of the sample treated by the vehicle control; as well as by ROS intensity that was 250% or above, or lipid intensity that was 200% or above the vehicle-treated control. The last “Test Score” column is a logical OR of the previous seven experimental results (i.e., a positive in any measurement would classify that drug as positive). Both DILI Label positives and Test Score positives were shown in bold. All these drugs were tested at 100-fold of the therapeutically active average single-dose plasma maximum concentrations ($C_{max}$) for 24 hours in sandwich-cultured primary human hepatocytes.
as potential mechanisms. Labetalol, an antihypertensive agent known to cause hepatotoxicity in some patients [94], increased oxidative stress and intracellular lipids in our human hepatocyte assay. Increased oxidative stress was also observed for chlorzoxazone and dantrolene, two muscle relaxants known to be associated with idiosyncratic but serious hepatotoxicity in the clinic [95]. In comparison, the bottom 13 drugs in Table 1.1 did not induce the same frequency of hepatotoxicity as the first 11 drugs, and they did not induce any significant changes in the human hepatocyte imaging assays. Among these 13 drugs were antidiabetic drugs pioglitazone and rosiglitazone, antidepressants amitriptyline and fluoxetine, NSAIDs aspirin and ketotifen, and antibiotics penicillin and paromomycin. These drugs have relatively safer clinical profiles with regard to hepatotoxicity compared to others within their respective therapeutic classes, as discussed earlier. The overall concordance of the human hepatocyte imaging assay technology (HIAT), when applied to over 300 drugs and chemicals, is about 75% with regard to clinical hepatotoxicity, with very few false-positives. The fact this relatively simple and high-throughput hepatocyte imaging assay can uncover so many common hepatotoxicity mechanisms previously reported in the literature made this technology especially attractive as a preclinical in vitro assay system to select drug candidates with improved TI for clinical hepatotoxicity.

1.4.3 Muscle Toxicity

Because of its high dependence on intracellular energy levels, muscle can be another major site of toxicity for pharmaceuticals [55]. As referenced in the beginning of this chapter, cerivastatin is a structurally distinct HMG-CoA reductase inhibitor. It effectively decreases LDL cholesterol at 1% to 3% of the doses and about 10% of the $C_{\text{max}}$ levels of previously available statins [96]. In preclinical safety evaluation in rats, mice, minipigs, dogs, and monkeys, cerivastatin exhibited a similar toxicologic profile to other statins and is well tolerated [97]. However, in postmarket adverse event reports (AERs) submitted to the US Food and Drug Administration (FDA), cerivastatin showed a higher propensity to cause muscle toxicity, especially when co-administered with fibric acid derivatives to lower both cholesterol and triglyceride levels [98]. This led to the market withdrawal of cerivastatin [99]. Reporting rates for all statins, except for cerivastatin, were similar and much lower than 1 per 100,000 prescriptions. The cerivastatin reporting rate was much higher at 4.24/100,000 prescriptions [98]. Could such a subtle, but important, difference be predicted by a therapeutic index screening approach, at least retrospectively?

The comparative cytotoxicity effects of a panel of statins on human skeleton muscle cells (HSkMCs) were studied in vitro [100]. These HSkMCs were derived from normal human fetal skeleton muscle, and were positive for sarcomeric myosin and were fused into multinucleated myotubes (Cell
Applications, Inc.). HSkMCs were cultured in 24-well culture plates and grown to semiconfluence. They were induced to differentiate by changing the growth medium to differentiation medium (Cell Applications, Inc.). Three days later, various concentrations of drugs were added and the cultures were incubated for 24 hours. Under the light microscope, cerivastatin induced significant cell damage in cultured HSkMCs at 0.1 to 1μM \([100]\), while other statins induced similar damage only at 10μM or higher. The cell damage induced by cerivastatin was abolished by the addition of mevalonolactone, suggesting the mechanism of toxicity may be linked to excessive antagonism of the HMG-CoA reductase activity in the HSkMCs. It was known that after a single oral 0.8 mg dose of cerivastatin, the \(C_{\text{max}}\) of total cerivastatin is about 8μg/L, or 0.016μM \([101]\). Thus, based on this limited HSkMC culture studies, a provisional safety margin of 30 times the efficacious \(C_{\text{max}}\) is needed to minimize the risk of muscle toxicity by cerivastatin and possibly other statin drugs. In several comparative studies among several statins, a consistent trend is that cerivastatin is the one with an exceptionally low TI in those in vitro test systems \([100,102–104]\).

### 1.4.4 Is Dose Escalation a Common Theme in Estimating Toxic Exposure?

As noted above, a provisional safety margin of 30- to 100-fold has been proposed for cardiac toxicity and hepatic toxicity respectively, and possibly 30-fold for myopathy. The requirement for an elevated exposure in the in vitro setting to identify deleterious organ effects of drugs may be due to a combination of (1) liver exposure to an orally dosed drug can be higher than its systemic exposure; (2) population PK variability due to age, genetics (including drug metabolism and transporters), and drug–drug interactions could further exacerbate local drug exposure and toxicity; (3) idiosyncratic organ history (including disease and previous drug exposures); (4) onset of toxicity in vivo is typically much longer than in vitro, thus requiring dose escalation in most short-term in vitro systems. Hence it is likely that dose escalation may be a common theme in estimating toxic exposure in the drug discovery stage. In the future, when all the variables in PK and pharmacodynamics can be accounted for and simulated a priori, a more precise estimation of TI may be possible.

Another theme is that the predictive safety margin of each in vitro system needs to be “calibrated” retrospectively using both toxic drugs and “clean” drugs in approaches similar to the examples given above. This is a prerequisite of building enough confidence that such predictions can be applied to new chemical entities for the same therapeutic target or targets. In the drug discovery stage, a TI prediction strategy needs to be practical to be applied broadly. Specifically at this stage, the volume of compounds that are generated and need to be evaluated typically exceeds several dozens to hundreds or thousands, the actual amount of compounds that are available for testing often
less than 100 mg, the time window for decision making is often limited to a few days to weeks. In these scenarios human and animal cellular systems cultured in well-characterized ways to recapitulate the relevant in vivo responses, coupled with high-throughput and high-content screening technologies, have the best potential for an efficient prediction of TI in the drug discovery stage.

1.5 FUTURE PERSPECTIVES

As illustrated in this chapter, the choices of cellular models, biomarkers being investigated, and the concentration–effect responses being measured have a profound effect on the therapeutic index prediction. Future efforts to improve TI prediction should therefore focus on these directions: (1) cellular models that better mimic the in vivo situation, (2) more accurate concentration measurement, and (3) systems based pharmacokinetic and pharmacodynamic prediction.

Although cellular models have been used in pharmacology and toxicology research for decades, it is only recently that the omics technology has been applied to the characterization of cell culture models [77,105,106] and the cell engineering field [107,108]. Primary cells cultured under defined conditions to better mimic the disease situation will be increasingly used as the result of better isolation, culturing, and characterization techniques. In particular, primary cells that maintain tissue-specific metabolic functions will be in increasing demand (primary neurons, beating cardiomyocytes, drug-metabolizing and polarized hepatocytes and nephrons, etc.). Stem cell technology promises to deliver unlimited sources of differentiated primary cells without the need to wait for donor availability [109,110]. Tissue engineering is an emerging field that can provide us with more physiological cell–cell interactions, endogenous functions, and well-defined fluid dynamics [111–113]. Taken together, these technologies promise to decrease the gaps between the in vitro and in vivo world.

However, technology alone cannot bring us closer to improved TI prediction without the sound principle of pharmacology and appropriate design of toxicological studies. Bioanalytical quantification of the actual “total” and “free” in vitro drug concentration in both whole cell efficacy and toxicity tests, and better PK prediction to project both efficacy and toxicity concentrations in vivo are still needed for any TI prediction. Recently human PK predictions based on in vitro systems have become more realistic [114,115]. This had led to the hope that in silico predictions of PK are not too far in the future [116–118]. Regarding pharmacodynamic predictions, mathematical simulations based on our understanding of human pathophysiology will become an integral part of any future drug R&D strategy [119–121]. With these technological advances, the science of projecting in vivo therapeutic index based on integrated in silico, in vitro, and in vivo approaches will flourish.
REFERENCES

FOCUS ON THE FUNDAMENTALS


REFERENCES

REFERENCES


REFERENCES


**ACKNOWLEDGMENT**

The authors extend profound gratitude to their past collaborators at Pfizer Global R&D, especially Margaret Dunn, Arthur Smith, Jeffrey Chabot, Peter Henstock, David de Graaf, Yue Chen, and Judith Treadway.