ADVANCING TOXICOLOGY-BASED CANCER RISK ASSESSMENT WITH INFORMATICS

Joel P. Bercu

Submitted to the faculty of the University Graduate School
in partial fulfillment of the requirements
for the degree
Doctor of Philosophy
in the School of Informatics,
Indiana University

May 2010

Accepted by the Faculty of Indiana University, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Malika Mahoui Ph.D., Chair

Pedro R. Romero Ph.D.

Doctoral Committee

James L. Stevens Ph.D.

April 16, 2010

Josette F. Jones Ph.D.

Mathew J. Palakal Ph.D.

ACKNOWLEDGEMENTS

I thank all of the coauthors for their work in the various chapters, especially Stuart Morton, J. Thom Deahl, Dr. Wherly Hoffman, Courtney Callis, Dr. Krista Dobo, Dr. Scott Obach, Robert Jolly, and Dr. James Stevens for their significant contributions to this dissertation. I also thank Autumn McClatchy and Stephanie Hines for their data entry, and Dr. Kirk Arvidson for providing us his classifications of structural alerts.

ABSTRACT

Joel P. Bercu

ADVANCING TOXICOLOGY-BASED CANCER RISK ASSESSMENT WITH INFORMATICS

Since exposure to carcinogens can occur in the environment from various point sources, cancer risk assessment attempts to define and limit potential exposure such that the risk of developing cancer is negligible. While cancer risk assessment is widely used with certain methodologies well accepted in the scientific literature and regulatory guidances, there are still gaps which increase uncertainties when assessing risk including: (1) mixtures of genotoxins, (2) genotoxic metabolites, and (3) nongenotoxic carcinogens. An *in silico* model was developed to predict the cancer risk of a genotoxin which improved methodology for a single compound and mixtures. Monte Carlo simulations performed with a carcinogenicity potency database to estimate the overall carcinogenic risk of a mixture of genotoxic compounds showed that structural similarity would not likely increase the overall cancer risk. A cancer risk model was developed for genotoxic metabolites using excretion material in both animals and humans to determine the probability not exceeding a 1 in 100,000 excess cancer risk. Two model nongenotoxic compounds (fenofibrate and methapyraline) were tested in short-term microarray studies to develop a framework for cancer risk assessment. It was determined that a threshold for potential key events could be derived using benchmark dose analysis in combination with well developed ontologies (Kegg/GO), which were at or below measured tumorigenic and precursor events. In conclusion, informatics was effective in advancing toxicology-based cancer risk assessment using databases and predictive techniques which fill critical gaps in its methodology.

Malika Mahoui Ph.D., Chair

TABLE OF CONTENTS

LIST OF ABBREVIATIONS viii

INTRODUCTION 1

Overview of Specific Aims 1

Background and Significance 2

Specific Aim 1 5

Specific Aim 2 7

Specific Aim 3 8

Overall Risk Assessment Framework 11

Figures 14

IN SILICO APPROACHES TO PREDICTING CANCER POTENCY FOR RISK ASSESSMENT OF GENOTOXIC IMPURITIES IN DRUG SUBSTANCES 15

Abstract 15

Introduction 15

Methods 17

Results 21

Discussion 25

Tables 28

Figures 34

QUANTITATIVE ASSESSMENT OF CUMULATIVE CARCINOGENIC RISK FOR MULTIPLE GENOTOXIC IMPURITIES IN A NEW DRUG SUBSTANCE 37

Abstract 37

Introduction 37

Methods 40

Results 47

Discussion 49

Conclusions 53

Tables 54

Figures 59

A STRATEGY FOR RISK ASSESSMENT OF HUMAN GENOTOXIC METABOLITES 64

Abstract 64

Introduction 65

Case Studies 77

Discussion 81

Tables 90

Figures 95

TOXICOGENOMICS AND CANCER RISK ASSESSMENT: A FRAMEWORK FOR KEY EVENT ANALYSIS AND DOSE-RESPONSE ASSESSMENT FOR NONGENOTOXIC CARCINOGENS 99

Abstract 99

Introduction 100

Methods 103

Results 106

Discussion 108

Conclusions 113

Tables 114

Figures 123

DISCUSSION 133

Advancement of Carcinogenicity Risk Assessment Framework 133

Specific Aim 1 134

Specific Aim 2 137

Specific Aim 3 140

Conclusions 143

Figures 145

REFERENCES 146

CURRICULUM VITAE

LIST OF ABBREVIATIONS

ADME	absorption, distribution, metabolism, excretion
NOAEL	no observable adverse effect level
LOAEL	lowest observed adverse effect level
TTC	threshold of toxicological concern
TD_{50}	average daily dose at which 50% of animals remains tumor free
AUC	area under the curve
HLA	human lymphocyte aberration assay
GERD	gastroesophageal reflux disease
ICH	International Committee on Harmonisation
IWGT	International Workshop of Genotoxicity Tests
CYP	cytochrome P450
HED	human equivalent dose
USEPA	United States Environmental Protection Agency
USFDA	United States Food and Drug Administration
POD	point of departure
CPDB	carcinogenicity potency database
GTI	genotoxic impurity
SAR	structure activity relationship
QSAR	quantitative structure activity relationship
BMD	the effective mean dose necessary to produce a 10% response
BMDL	the lower 95th percent confidence interval of a 10% response on a dose-response curve
GO	gene ontology
ALARP	as low as reasonably practicable

INTRODUCTION

Overview of Specific Aims

Cancer is a devastating disease and exposure to carcinogens through food, water, air etc. contributes to the overall cancer burden. Since eliminating all carcinogens from the environment is impossible, it is essential to quantify the potential risks and focus on those situations where reducing exposure can protect human health. Unfortunately current risk assessment methods do not address exposure to mixtures or metabolites, which are common, and public pressure to reduce animal testing prevent accumulation of sufficient animal data to perform a more traditional risk assessment.

Informatics and gene expression technologies offer solutions to gaps in carcinogenicity risk assessment methodology. We proposed to use these technologies to identify the threshold of toxicological concern (TTC), i.e. the dose at which the excess cancer risk for humans is negligible, for mixtures of chemicals and metabolites. The TTC concept is a common risk assessment tool that is currently applied to food, water, pharmaceuticals, and personal care products for exposure to low levels of genotoxic substances. **The overall goal of this research was to develop 3 new computational approaches to identify a TTC that can be used in a carcinogenicity risk assessment framework when limited data are available.** These approaches are novel changes as there are no scientifically-based risk assessment strategies for these potential human carcinogens. Computational strategies can provide leverage to deal with the limitations raised by lack of data in these cases.

Although the use of a TTC is common for single chemicals, the approach has not been applied to mixtures. The first goal was to provide risk assessment methodology for mixtures of genotoxic compounds. **In Aim 1, a carcinogenicity potency database (CPDB), was analyzed to determine the risks of multiple genotoxic compounds by establishing the TTC for genotoxic mixtures.**

Many chemicals are metabolized to excretable products, but there is no risk assessment methodology available to determine the carcinogenic risk when metabolites are genotoxic. The second goal was to provide a risk assessment framework for establishing provisional thresholds for exposure to genotoxic metabolites. Some suggest that the TTC concept is an appropriate method for assessing metabolite risk, but the hypothesis has not been tested. **In Aim 2, the CPDB was used to develop a model to determine the carcinogenic risk of a compound if its metabolite is genotoxic.**

Proving that a threshold exists has a dramatic impact in the risk assessment of a chemical. However, current approaches require extensive mechanistic analysis in order to establish a threshold. The final goal was to determine if global gene expression data from short-term microarray studies and systems biology analyses can be used to establish effect- thresholds, as a surrogate for the TTC, for nongenotoxic carcinogens to develop a permissible daily exposure in lieu of a more conservative approach. **In Aim 3, a gene expression analysis of nongenotoxic carcinogens was used to determine if mining gene expression data can establish existence of an effect-threshold.**

This chapter will provide background information and discuss the importance of Specific Aims 1, 2, and 3. The subsequent chapters (2-5) will provide the studies which satisfy the Specific Aims. The final chapter (Chapter 6) will provide some overall discussion of the Aims in context with the data presented in the previous chapters.

Background and Significance

Cancer risk assessment

Cancer risk assessment is a critical part of public health because it allows the effective control of potentially dangerous chemicals in the environment. While extensive efforts have been made over the years, there are still many areas that need improvement especially when gaps are observed in the toxicology data. *The purposes*

of Aim 1, 2, and 3 were to advance current cancer risk assessment techniques specifically for those instances where toxicology data is limited.

Overview of current cancer risk assessment methodology

Cancer is a devastating disease and is either not preventable (e.g. genetic predisposition), part of life-style choices made by the individual (e.g. smoking, diet, etc.), or as a result of environmental exposure to a carcinogen. Environmental exposure has been the source of regulatory attention because exposure can be controlled and is involuntary. Risk assessment is a process in toxicology which helps determine what level of exposure is considered safe (BEST 2008). Risk assessment contains fundamental elements which aids in the process including 1) hazard identification, 2) exposure assessment, 3) dose-response and 4) risk characterization.

Hazard identification is the process of identifying potential toxic effects in humans that can occur at any dose. Carcinogenicity hazards can be understood from a variety of assays; genotoxicity is one of the most common parameters measured. Genotoxicity studies can be predictors of carcinogenicity and are based on the ability of the compound or its metabolites to interact with the DNA to cause mutations, or chromosomal aberrations (ICH 1996, 1997; USEPA 1986). Even though genotoxicity is a precursor to carcinogenicity, not all carcinogens are genotoxic and not all genotoxic compounds are carcinogenic (Ames *et al.* 1993; Kim and Margolin 1999; Kirkland *et al.* 2005; Kirkland *et al.* 2006; Zeiger *et al.* 1996). Carcinogenicity can be secondary to nongenotoxic toxicity such oxidative stress or cellular proliferation (Klaunig *et al.* 2000). Also, there are many repair processes in the body that can prevent the formation of tumors from a genotoxic insult (Ames *et al.* 1993; Doak *et al.* 2007; Jenkins *et al.* 2005). Carcinogenicity studies can be performed in animals, typically for the lifespan of the animal (Haseman 2000). Epidemiology studies in humans can determine if a compound is a known human carcinogen. Typically, carcinogenicity and epidemiology studies are

very resource intensive. Therefore in many cases the only carcinogenicity hazard that is known for a compound is its genotoxicity.

Exposure assessment involves the quantification of exposure that a person might receive. Trace levels of carcinogens may exist in the environment such as food, water, or air, but detection depends on the sensitivity of the analytical limit (Kroes *et al.* 2004). If levels observed in the environment are determined to be unacceptable from a toxicology perspective then strategies are developed to reduce the exposure (USEPA 2000a).

Dose-response assessment involves understanding the potential risks from available toxicology information (USEPA 2002). Animal data is typically applied since controlled experiments with various doses are available. The dose-response assessment for carcinogenicity depends on if the chemical exhibits a threshold (Clewell 2005; Conolly 1995; USEPA 2005a). Typically if the compound is a genotoxic carcinogen, then it is assumed to have no threshold (Barlow *et al.* 2006; USEPA 2005a). A threshold means that there is a dose where higher exposures have an effect and lower exposures do not. However if there is no threshold, then an effect can occur at any dose. Therefore, a linear extrapolation is performed from animal studies to determine the risk of developing cancer over background from chemical exposure. By nature, there is a risk from any given exposure no matter how low.

Risk characterization is the final step and involves comparing the dose-response assessment to exposure to determine the risk of developing cancer from the compound (USEPA 2000a). Once this characterization is made, decisions are facilitated from the information provided. Every effort is made to reduce exposure to as low as reasonably practicable (ALARP) (CHMP 2006). Risk characterization allows organizations to make decisions and maximize their resources while ensuring safety. Our goal was to provide

adequate methodology for risk characterization for mixtures and metabolites that have genotoxic potential.

Assessing risk based on limited data - Threshold of toxicological concern

Ideally, sufficient data are available to quantify risks with little uncertainty. Unfortunately, with the volume of chemicals and demand to limit animal testing toxicity testing for chemicals can be limited. Nonetheless, risk assessment processes have been developed to evaluate compounds with limited data (Dourson 1996; Dourson *et al.* 1996; Dourson and Stara 1983; Dybing *et al.* 2002; Fiori and Meyerhoff 2002; Kroes *et al.* 2005). Typically the assessments are conservative to account for the uncertainties that exist in the data. The threshold of toxicological concern (TTC) is typically applied to compounds were limited data exists for a molecule (Kroes *et al.* 2005; Kroes and Kozianowski 2002; Kroes *et al.* 2004; Renwick 2005). The TTC can be derived based on structural information about the molecule or predictive toxicity tests such as cellular-based assays. While the TTC may be limited in its precision to predict risk, it is conservative and protects for uncertainty that exists with limited data. While the TTC has been applied in many critical areas, there are still many areas in which it may add value (Blackburn *et al.* 2005; Dolan *et al.* 2005; Kroes *et al.* 2004; Müller *et al.* 2006).

Specific Aim 1

In many cases genotoxicity testing is only available to understand carcinogenic hazards of a compound. The TTC is the current accepted methodology for analyzing the risk of a single genotoxic compound, but complexities exist with mixtures of genotoxic compounds. *The purpose of Aim 1 was to use the TTC to understand the total carcinogenic risk when exposed to a mixture of genotoxic compounds.*

TTC - Risk current approaches to genotoxic compounds

The TTC has been applied by others when genotoxicity but not carcinogenicity is known about a compound (CHMP 2006; Kroes *et al.* 2004; Müller *et al.* 2006). It was

developed from a database of known carcinogens (cancer potency database - CPDB) (Gold *et al.* 1999; Gold *et al.* 2005; Gold *et al.* 1984; Gold *et al.* 1991). For each carcinogen there was an associative TD_{50} value, the dose at which 50% of animals remain tumor free over background when exposed over a lifetime (Peto *et al.* 1984; Sawyer *et al.* 1984). In context, the TD_{50} is an effective measure of carcinogenic potency. Based on an analysis of the CPDB, a TTC was developed, which represents a low probability of exceeding a negligible excess cancer risk (Kroes *et al.* 2004; Müller *et al.* 2006; Munro *et al.* 1999).

Scientific gap - Limitations of the TTC approach for mixtures

The TTC concept was developed to assess risk for exposure to an individual compound. However, in the environment individuals are exposed to mixtures of compounds. Currently there is no methodology available for assessing the risk of multiple genotoxic compounds. This has been a scientific gap for pharmaceuticals where an individual can be exposed to multiple genotoxic impurities. The regulatory guidance for genotoxic impurities in pharmaceutical products states that structural similarity is the critical for assessing the cancer risk (CHMP 2006). The philosophy is that structurally similar compounds act similarly from a mechanistic perspective and therefore have a combined limit that does not exceed the TTC. Structurally dissimilar compounds are thought to act through independent mechanisms, thus allowing for separate limits. Although useful as a guidance, this conservative regulatory perspective is not based on a scientific method. Therefore, research is needed to determine the cumulative cancer risk of multiple genotoxic compounds. While this has been a particular concern for pharmaceuticals, this essentially impacts all applications of the TTC. In Aim 1, we developed methodology for the TTC concept to be applied in a risk assessment framework for mixtures of genotoxic compounds.

Specific Aim 2

Many chemicals are metabolized to genotoxic species. Thus, elucidating the fate and risk for potentially genotoxic metabolites is critical to understanding the safety of a molecule. While extensive methodology has been developed to understand metabolism, methods to assess the risk of metabolites with genotoxicity concerns is not well developed. Assessing the risk of genotoxic metabolites is essential to understand the safety from exposure to the compound and its metabolite. *The purpose of Aim 2 was to develop methodology for understanding the carcinogenic risk following exposure to a genotoxic metabolite and the utility of current risk assessment methods, such as the TTC.*

Metabolism background

Many foreign compounds when ingested undergo extensive biotransformation, yielding a variety of systemic metabolites. Safety testing for a compound requires an understanding of the potential toxicity of these metabolites (Baillie *et al.* 2002; Luffer-Atlas 2008; Smith and Obach 2006; USFDA 2008b). There are many examples of carcinogens that require metabolism to genotoxic species in order to produce tumors (Ku *et al.* 2007). Thus, differences in metabolism between species and individuals can affect the carcinogenicity of a compound. This creates challenges when assessing the safety of a compound because its metabolic fate in humans can differ from the testing model. Furthermore, this can be an issue from a population perspective where certain individuals could be sensitive to the exposure of a compound because of their unique metabolic enzymes.

Scientific gap - No framework for risk assessment of genotoxic metabolites

There have been efforts to develop a common framework to manage the safety of metabolites, especially in pharmaceuticals (Baillie *et al.* 2002; USFDA 2008b). Special considerations warrants the development of a framework for assessing the

safety of potentially genotoxic metabolites given that the importance of metabolism in carcinogenicity (Ku *et al.* 2007). Furthermore, the ability of both *in vitro* and *in vivo* genotoxicity test systems to model human metabolism can be variable requiring careful scrutiny of these assays (Obach and Dobo 2008). Despite the need for evaluating genetic safety of metabolites, no practical recommendations are available to address safety of identified human genotoxic metabolites. Thus, the methods developed under Aim 2 will be a significant contribution to risk assessment methodology for xenobiotic metabolites.

Developing the TTC for metabolites

A working group which convened at the 4th International Workshop of Genotoxicity Tests (IWGT) acknowledged the need for a practical strategy to respond to documented human metabolite exposures and suggested that the TTC may be used to support a risk assessment approach for genotoxic human metabolites (Ku *et al.* 2007). However, the practical utility and quantification of risk has not been determined to test the hypothesis that the TTC is an appropriate endpoint.

Metabolites are unique in that their systemic exposure in animals is typically measured and the risk to the parent may be better understood from exposure to the metabolite. Systemic exposure to metabolites can differ between humans and animals due to species differences. Therefore, a model may be developed that takes into account rates of metabolism in applying the TTC concept which would improve its application to risk assessment of genotoxic metabolites. In Aim 2, we extended the TTC concept to metabolites by establishing a model based on human and animal exposure information.

Specific Aim 3

Since cancer is such a devastating disease and effects at low doses are unknown, conservatism is typically the default practice for carcinogens. This leads to

regulations for exposure that have broad economic and human health consequences. The typical approach for a genotoxic carcinogen is to assume that the effect is linear and cancer can occur from even a small exposure (USEPA 2005a). However, for nongenotoxic carcinogens there may be an opportunity to revise the assumption so that a threshold approach can be used in lieu of the more conservative assessment (Butterworth 2006; Melnick *et al.* 1996). This assumption is important, because it means that one can raise the limit for nongenotoxic carcinogens several orders of magnitude. However, being nongenotoxic is not sufficient to use a less stringent approach, and one must show the existence of a threshold. *The purpose of Aim 3 was to develop methodology that can be used to determine a safe level for nongenotoxic carcinogens based on short-term toxicogenomic information.*

Establishing the threshold based on gene responses

Determining dose-response at a gene transcription level is challenging given the high dimensionality of the data set; thus, dose-response analysis is typically performed at the whole-organism level. Genomic analysis enriches our knowledge-base, however, because biological responses can be understood at a mechanistic level versus gross pathology. Recent advancement in informatics, has allowed analysis of dose-response at a gene transcription level (Yang *et al.* 2007; Yu *et al.* 2006). Genomic analysis of formaldehyde, allowed for a description of different toxicity-associated mechanisms and understanding dose-response for these mechanisms (Andersen *et al.* 2008; Thomas *et al.* 2007). In Aim 3, we used transcript profiles to establish dose-responses at a gene level to better understand dose-dependence for nongenotoxic carcinogens.

Mode of action - essential for establishment of a threshold for nongenotoxic carcinogens

Under current regulatory guidance from the USEPA, a mode of action (MOA) has to be described for a compound so that a threshold can be demonstrated (USEPA 2005a). The MOA is a description of a sequence of events that results in the formation

of tumors in animals. The MOA must have one or more key event(s), which is a biological precursor event that is necessary to cause cancer. The MOA differs from mechanism which requires a more detailed understanding of the toxicity. For example, for compounds that act as ligands for the peroxisome proliferator activator receptor (PPAR), the MOA is known to involve a key event which is induction of peroxisome proliferation (Yu *et al.* 2003). Although the mechanism that leads from peroxisome proliferation to cell transformation and tumor induction are not known, knowledge that peroxisome proliferation is required has been sufficient to establish that drugs (fibrates) that are PPAR activators are rodent specific carcinogens (Klaunig *et al.* 2003) clearing the way for clinical application. Unfortunately, when a MOA and key event cannot be identified, then the conservative linear approach is then applied as a default. A key challenge for using a threshold approach for nongenotoxic carcinogens has been identifying these key events to allow a MOA argument to be applied. Although detailed mechanistic analysis is not required per se, practically it is often a requirement using traditional toxicology studies so that a key event can be identified.

Systems biology using microarray analysis provides a unique opportunity to analyze potential key events. Ontologies such as Kegg pathways (Kanehisa and Goto 2000; Kanehisa *et al.* 2006) or Gene Ontology (GO) (Ashburner *et al.* 2000) represent our knowledge of biological processes and pathways; therefore, analysis of these pathways (Dennis *et al.* 2003; Huang da *et al.* 2009) can reduce dimensionality of data and represent a majority of potential key events. It is important to note that the USEPA, does not require all key events to be identified, just key event(s) that are linked to a dose-response and establish a threshold (USEPA 2005a). Therefore, in Aim 3, we established methods to analyze potential key events using standard ontologies and linked them with potential thresholds using dose-response methodology.

Overall Risk Assessment Framework

Although the Specific Aims deal with different aspects of risk assessment for carcinogens, e.g. genotoxic vs. nongenotoxic agents and single agents vs. mixtures, they are interconnected as part of an overall decision framework designed to advance cancer risk assessment. Figure 1 demonstrates the current carcinogenicity risk assessment framework with gaps addressed by the three Specific Aims. If a compound lacks carcinogenicity information, then genotoxicity data for the compound are used to assess risk. For nongenotoxic compounds without carcinogenicity information, no carcinogenicity risk assessment is required. The TTC is typically applied for genotoxic compounds. If there is a mixture of genotoxic compounds, then an additive limit is applied for structurally similar compounds while separate limits are applied for dissimilar compounds. There are limitations to this approach mainly because the TTC is a conservative default that can be overly restrictive in some cases and the mixtures assessment is based on judgement and not science. The focus of Aim 1 was to improve upon these limitations.

If there is carcinogenicity data available and the compound is not carcinogenic, genotoxic metabolites may also be considered. Exposure to the genotoxic metabolite must be high enough in animals for the bioassay to be an adequate model for tumorigenicity assessment, or exposure in humans must be negligible. If no genotoxic metabolites are identified, then no carcinogenicity risk assessment is required. However, if a metabolite is genotoxic or is predicted to be genotoxic, there is currently no methodology to assess the risk of identified genotoxic metabolites. Aim 2 provided an overall framework for understanding the risk of genotoxic metabolites using results from the cancer bioassay and exposure in humans. For genotoxic carcinogens, it is assumed that there is no threshold and linear low-dose extrapolations are applied.

Under current guidelines, risk assessment for nongenotoxic carcinogens may allow for a less conservative approach; however, a threshold-based MOA must be identified. The current methodology for identifying a threshold is limited and depends in large part on the ability to identify 'key events' linked to the mechanism of carcinogenesis. Determining mechanism is difficult and time consuming. In Aim 3, a toxicogenomic approach is described that advances the methodology by facilitating the identification of key events and establishing a threshold-based MOA.

In order to perform an adequate risk assessment for any agent, the decisions outlined in Figure 1 must be considered. Gaps in knowledge are as highlighted in Figure 1 will be addressed using computational methods. Each Specific Aim outlined in the thesis is crtical for developing methodology for an overall risk assessment framework. Therefore, all of the Aims are interdependent; if one part of the framework fails, then the overall framework fails to correctly characterize risk. The focus for each Aim was to improve a part of the framework where critical technical gaps were identified. As each part of the framework is strengthened, the uncertainties surrounding cancer risk assessment is reduced.

<u>Reliable methods for risk assessment are critical to public health</u>

Chemicals have benefit within society, but the use of chemicals such as drugs or agricultural products are also associated with risks. Toxicology risk assessment, allows for the understanding of risks following exposure to a chemical. Given the risk/benefit equation, it is important that risk assessment strive to be as accurate as possible. If the assessment is too conservative from overestimation of risk, then resources are spent on action plans that would not significantly improve public health; thus expending resources which could be better used in other areas. However, if the assessment underestimates the risk, then decisions are made that can compromise safety. The reality is that decisions are made even when toxicology data is missing. In this proposal we will

develop methods for cancer risk assessment that can be used to improve its accuracy even for incomplete datasets. Some critical gaps exist for mixtures of genotoxic compounds, genotoxic metabolites, and nongenotoxic carcinogens. Therefore, the focus was to improve upon these methodologies and build upon the current risk assessment strategies for carcinogens.

Figures

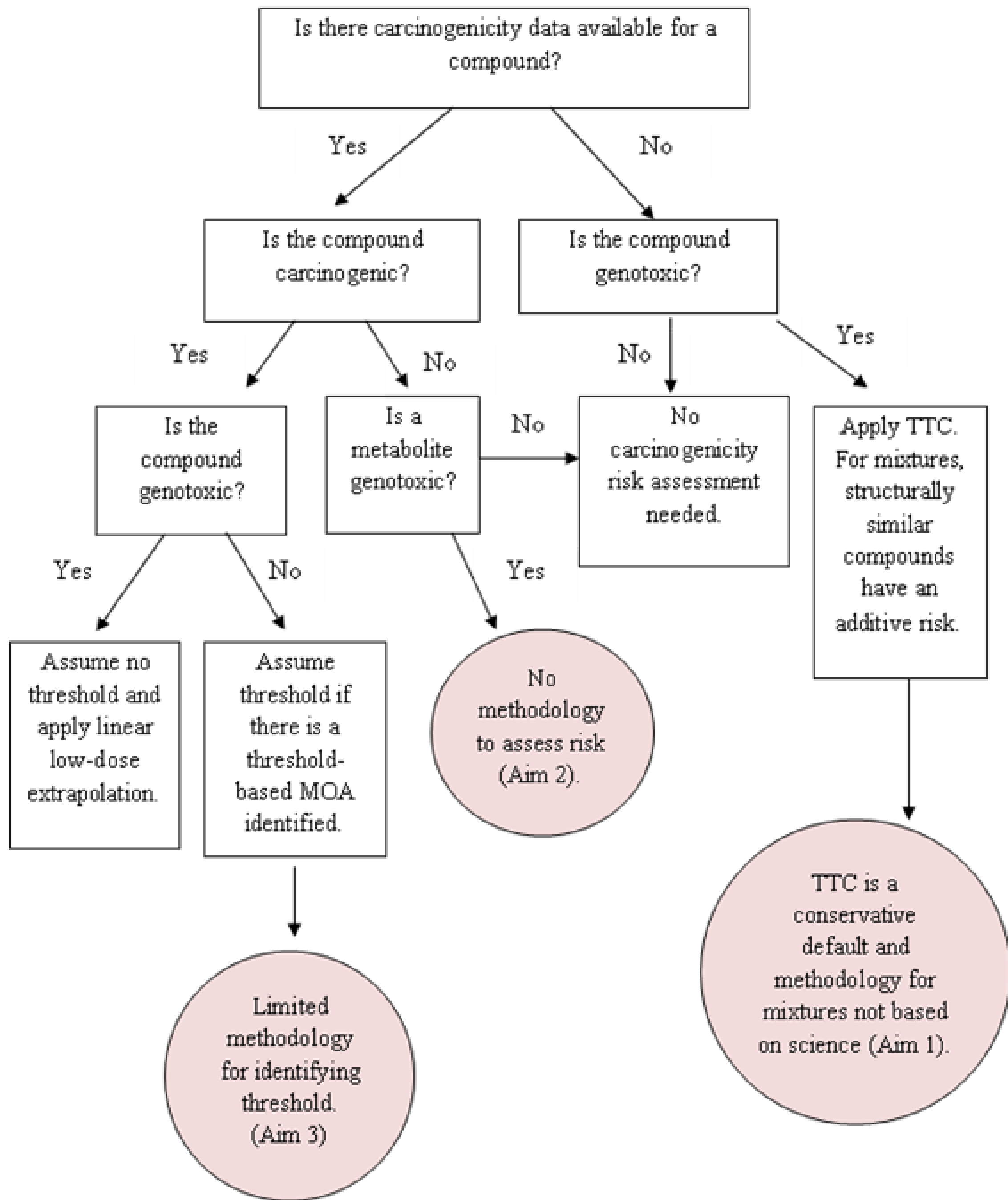


Figure 1. Overall framework for cancer risk assessment and gaps addressed by Specific Aims. Gaps in current methodology are shown in circles with the Specific Aim designed to address that gap noted.

IN SILICO APPROACHES TO PREDICTING CANCER POTENCY FOR RISK ASSESSMENT OF GENOTOXIC IMPURITIES IN DRUG SUBSTANCES[1]

Abstract

The current risk assessment approach for addressing the safety of very small concentrations of genotoxic impurities (GTIs) in drug substances is the threshold of toxicological concern (TTC). The TTC is based on several conservative assumptions because of the uncertainty associated with deriving an excess cancer risk when no carcinogenicity data are available for the impurity. It is a default approach derived from a distribution of carcinogens and does not take into account the properties of a specific chemical. The purpose of the study was to use *in silico* tools to predict the cancer potency (TD_{50}) of a compound based on its structure. Structure activity relationship (SAR) models (classification / regression) were developed from the carcinogenicity potency database using MultiCASE and VISDOM. The MultiCASE classification models allowed the prediction of carcinogenic potency class, while the VISDOM regression models predicted a numerical TD_{50}. A step-wise approach is proposed to calculate predicted numerical TD_{50} values for compounds categorized as not potent. This approach for non-potent compounds can be used to establish safe levels greater than the TTC for genotoxic impurities in a drug substance.

Introduction

Control of impurities is an important part of the drug manufacturing process as it improves quality and minimizes safety risks. The ICH Q3A(R2) and Q3B(R2) guidances recommend limits for reporting, identifying and qualifying drug substance and drug product impurities (ICH 2006a, b). The guidance recognizes that lower limits may be appropriate for impurities considered unusually toxic. Recent guidances exist for a

---

[1] This chapter has been published previously in: Bercu JP, Morton SM, Deahl JT, Gombar VK, Callis CM, and van Lier RBL. (2010). In silico approaches to predicting cancer potency for risk assessment of genotoxic impurities in drug substances. Regul. Toxicol. Pharmacol. (In Press).

special class of toxic impurities, which are considered genotoxic (CHMP 2006; USFDA 2008a). An impurity may be classified as genotoxic if it is positive in the Ames mutagenicity assay or predicted to be positive in the assay based on structure (Dobo *et al.* 2006). The presumption for genotoxic impurities (GTIs) is that they are also carcinogens with no threshold unless proven otherwise (Barlow *et al.* 2006; CHMP 2006; USEPA 2005a).

The level of a GTI must be as low as reasonably practicable (ALARP) and below a limit considered toxicologically acceptable. The current risk assessment paradigm for addressing the safety of GTIs, with unknown carcinogenic potency, is the threshold of toxicological concern (TTC) (Kroes *et al.* 2004; Müller *et al.* 2006). The TTC is a dose at which there is a high probability of being below a negligible excess cancer risk (Kroes *et al.* 2004; Munro *et al.* 1999). The accepted negligible cancer risk for GTIs in human pharmaceuticals of 1 in 100,000 is associated with an exposure of 1.5 µg/day (CHMP 2006; USFDA 2008a). The TTC was developed from a database of rodent carcinogens with TD_{50} values as the measures of carcinogenic potency (Gold *et al.* 1999; Gold *et al.* 2005; Gold *et al.* 1984; Gold *et al.* 1991; Gold *et al.* 1992). The TD_{50} is defined as the daily dose in mg/kg/day at which there is a fifty percent probability of developing tumors over background (Peto *et al.* 1984; Sawyer *et al.* 1984).

The TTC is based on several conservative assumptions outlined previously because of the uncertainty of estimating a cancer potency value from sparse data (Delaney 2007; Humfrey 2007; Krewski *et al.* 1990). Delaney (2007) theorized that a higher threshold than the TTC likely exists for compounds used in the synthesis of a drug substance because their structures are not part of the “potent class” (Delaney 2007). However, there are no known methods to determine purely on structure whether a compound would be of higher carcinogenic potency.

In silico models have been shown to be effective in predicting a number of toxic endpoints such as potential genotoxicity or carcinogenicity (Hayashi *et al.* 2005; Matthews and Contrera 1998, 2007; Matthews *et al.* 2006b; Mayer *et al.* 2008; Votano *et al.* 2004; White *et al.* 2003). There have been some recent efforts to develop a quantitative structure activity relationship (QSAR) of cancer potency using toxicity estimates and chemical properties (Venkatapathy *et al.* 2009). The toxicity estimates and chemical properties were either experimentally derived or computed. Limited predictivity was observed for a single estimate, with an improvement occurring after applying classification and a regression tree (Venkatapathy *et al.* 2009).

In many cases, risk assessment of an impurity is based only on its structure and limited experimental data (e.g. Ames bacterial mutation assay results). Therefore, the typical risk assessment approach defaults to the TTC. However, a recent draft guidance developed by the USFDA for GTIs allows for higher levels than the default TTC. The guidance suggests calculating cancer risk based on carcinogenic potency from a structurally similar known carcinogen (USFDA 2008a) and adopting that calculation for the GTI limit. This offers an opportunity to employ *in silico* models to predict carcinogenic potency (TD_{50}) solely from the chemical structure and provide a more robust risk assessment.

The purpose of this study was to utilize *in silico* methodologies for predicting the carcinogenic potency to improve the risk assessment of GTIs. This approach may provide the opportunity to make more practical decisions surrounding GTIs versus default assumptions, and ultimately, enhance the drug manufacturing process.

Methods

Two software tools were applied to derive cancer potency predictions. The first was MultiCASE from MultiCASE, Inc. (www.MultiCASE.com) and the other was an in-house package, VISDOM. MultiCASE, also called MC4PC (v.2.1) for use with MS

Windows software, is a machine-learning application that has been validated and used by the USFDA (http://www.fda.gov/AboutFDA/CentersOffices/CDER/ucm092125.htm) for prediction of toxicity (Benz 2007; Matthews and Contrera 2007; Matthews *et al.* 2008; Matthews *et al.* 2009). MultiCASE breaks down all input chemical structures into non-cyclic fragments of two to ten atoms. The experimental toxicity value (TD_{50} in this case) of a compound is translated into "CASE units" ranging from 10 to 90 and compounds are categorized into three classes based on the assigned "CASE units". Compounds with CASE units between 10 and 19 are labeled inactive (or non-toxic), compounds with CASE units between 30 and 90 are called active (or toxic), and compounds with CASE units between 20 and 29 are grouped as "Marginal" (Gombar *et al.* 2007; Klopman 1984, 1992). Based on the label assigned to each compound, all fragments of that compound are tagged with that label. The probability of association between each fragment and the activity class is then established. The fragments that have statistically significant association with the "active" class are called "biophores" and fragments that have statistically significant association with the "inactive" class are called "biophobes". Depending on the biophores and/or biophobes present in a molecule, MultiCASE computes the probability of the molecule being associated with the active class. When a molecule is predicted to be active, MultiCASE also calculates a quantitative value of activity in CASE units.

VISDOM is an in-house package developed for deployment of predictive models at Eli Lilly and Company (Gombar and Zhang 2007). The model development techniques of VISDOM are similar to those in the commercial toxicity prediction system, TOPKAT (Gombar and Enslein 1995). Along with a TD_{50} value, the model predicts a 95% confidence interval around the prediction and a number of diagnostic measures to assess the reliability of the prediction. As in TOPKAT, VISDOM also has a reverse

QSAR (rQSAR) functionality to quantify the contribution of every non-hydrogen atom into the predicted TD_{50} value.

The TD_{50} predictor installed in VISDOM is a multiple linear regression QSAR model. The model employs information-rich structure quantifiers expressing electrotopological state (ES) values (Hall *et al.* 1991). ES values are computed on the constituent isolated and bonded hydrides in a molecule (Gombar *et al.* 2004). The sum of atomic ES values is used as a measure of molecular bulk. The ability of a molecule to form inter- and intra-molecular hydrogen bonds was quantified by the sum of ES values on hydrogen bond acceptors and donors (Lipinski 2000). ES descriptors simultaneously capture information about the arrangement and accessibility of electrons. Simply put, the ES of an atom is context-specific state and is computed from the number of sigma, pi, and lone pair electrons, and the number of bonded hydrogens, with the context specificity being derived from the number and arrangement of other atoms in the molecule.

<u>Data Set</u>

This modeling effort was focused on categorizing and estimating the potential cancer potency of individual organic compounds. The TD_{50} data used for the present work were extracted from the carcinogenicity potency database (CPDB) (http://potency.berkeley.edu/). From a total of 1515 compounds downloaded, a set of 694 was retained; noncarcinogenic compounds, mixtures, inorganics, and structural duplicates were removed from the dataset. Each TD_{50} in mg/kg/day was converted to a pTD_{50} for data normalization by Equation 1.

$$pTD_{50} = -\log\left(\frac{TD_{50}}{1000 \text{ x Molecular Weight}}\right) \quad (1)$$

Dividing by molecular weight transforms the cancer potency value on a molar basis. Histograms indicate that the pTD_{50} datasets were normally distributed, an advantage for statistical analyses, while the TD_{50} datasets were skewed (Figure 1).

Separate datasets contained TD_{50}s of rats and mice. The most potent TD_{50} from all target organ sites was developed for each species. If more than one TD_{50} was developed for an organ site and species, then the harmonic mean of the TD_{50}s was applied. This is consistent with the approaches used to develop the CPDB (http://potency.berkeley.edu/td50harmonicmean.html).

TD_{50} values were categorized as *potent* or *not potent*, or as falling in between these two categories and considered to be *indeterminate*. The criteria for the potent category was selected to be consistent with the value associated with the TTC for genotoxic compounds, an exposure value of 1.5 µg/day for a 70 kg person (Müller *et al.* 2006). This 1.5 µg/day value, representing an excess cancer risk of 10^{-5}, was linearly extrapolated to a TD_{50} value of 1 mg/kg/day and associated with a pTD_{50} of 4.53, the lower bound for all molecular weights in the set of compounds with TD_{50} values ≤ 1 mg/kg/day. So pTD_{50} values ≥ 4.53 were considered to be potent. In order to clearly separate the potent and not potent categories, the not potent category was selected to start at a TD_{50} level ten times higher than the potent category. The TD_{50} values for the not potent category were ≥ 10 mg/kg/day. The TD_{50} value of 10 mg/kg/day was associated with a pTD_{50} of 3.75, the lower bound for all molecular weights in the set of compounds with TD_{50} values ≥ 10 mg/kg/day. The distribution of the carcinogenicity data is shown in Table 1. Since the pTD_{50} cutoff was based on the lower bound for all molecular weights, many compounds that would have been categorized as not potent using TD_{50}s were labeled potent using pTD_{50}s. This reflects conservatism in the model because more compounds were labeled to be potent than exist below the TTC, which was derived from TD_{50}s. A TD_{50} prediction from MultiCASE or VISDOM was considered

invalid when either the prediction diagnostic failed, or if MultiCASE had a “Marginal” prediction as defined by its diagnostics.

Results

Model Development

Having observed differences in carcinogenic potency of compounds in rats and mice, the data set was separated to develop species-specific TD_{50} predictors. In order to test our TD_{50} models, 10% of the compounds were evenly selected from the TD_{50}, potency-ordered data for both rat and mouse and set aside as test sets. The training set consisted of the rest (90%) of the compounds. In order to provide accurate structural characterization for each compound in the model, canonical SMILES were coded for each chemical from PubChem (http://pubchem.ncbi.nlm.nih.gov/). MultiCASE and VISDOM models were developed using the training set data.

Both MultiCASE and VISDOM models were developed to minimize the effects of over-fitted models, inclusion of model-influencing compounds and observation-sensitive variables, and use of redundant and highly correlated descriptors. The quality and robustness of the models were determined through statistical parameters such as sensitivity, specificity, concordance, applicability, positive prediction value (PPV), and negative prediction value (NPV) in both resubstitution and cross-validation tests.

A consensus prediction approach was applied to the predictions of both MultiCASE and VISDOM models. There were several rules applied to the consensus model. If either model predicted the TD_{50} to be potent then the consensus was potent. If both models predicted the TD_{50} to be not potent then the consensus was not potent. If either model could not make a prediction then the result was defaulted to the result of the other model. Finally, if both models could not make a prediction, the consensus model result was removed from the analysis.

Additionally, since VISDOM provided a numerical prediction of a $p\text{TD}_{50,}$ statistics such as R^2 and adjusted R^2 determined the model's fit and Q^2 determined the model's performance in cross-validation and test sets (Golbraikh and Tropsha 2002). R^2 and Q^2 calculations were performed using JMP (v 6.0, SAS Institute). These predicted $p\text{TD}_{50}$ values were then converted back to TD_{50}s for the practical applications in risk assessment.

Validation

The validation statistics indicate that the models developed were conservative in that compounds are predicted potent more often than not (Table 2). Overall, excluding the indeterminate compounds from the test set resulted in improved statistics, especially sensitivity, which is the percent of experimentally potent compounds, predicted to be potent. When excluding test set indeterminate compounds, sensitivity of each model was relatively high: 94% (MultiCASE-Rat), 86% (MultiCASE-Mouse), 85% (VISDOM-Rat), 60% (VISDOM Mouse), 86% (Consensus-Rat) and 88% (Consensus-Mouse). The specificity, which is the percent of experimentally not potent compounds predicted to be not potent, was not as high in some areas as sensitivity reflecting the intended conservatism of the model where in some instances there were not potent compounds predicted potent. There was a high overall specificity in mouse models (81% MultiCASE, 100% VISDOM, and 85% Consensus) and individually the rat models were low (50% MultiCASE, 44% VISDOM, and 36% Consensus). Concordance was high in each rat and mouse model (68-90%), even with the low specificity. The amount of confident predictions improved with the consensus models (63-70% applicability versus 43-58% for the individual models). The explanation is that in some instances one model could make a prediction for a compound when the other was not confident.

Table 3 shows the model fit and resubstitution statistics for the VISDOM rat and mouse models. Essentially the models developed were robust with R^2 values ranging

from 0.66-0.79. This table indicates that the ability to predict precise values was better in the mouse model (Q^2 = 0.60) than the rat model (Q^2 = 0.34).

Following conversion back from the pTD$_{50}$, each predicted TD$_{50}$ was divided by the experimental TD$_{50}$. Therefore, higher numbers represent the undesirable situations (non-conservative) when the numerical VISDOM prediction was less potent than the experimental value. Table 4 represents the percent of compounds where the ratio was less than or equal to 1, 2, 5, and 10-fold the experimental result. A majority of compounds (rat: 86%; mouse:88%) had VISDOM TD$_{50}$ predictions that were ≤ 5-fold the experimental value. In addition, it was rare for VISDOM TD$_{50}$s to exceed 10-fold the original value, with 86% or 97% less than or equal to 10-fold the experimental values. Out of the four compounds that were greater than 10-fold from the original prediction, two were predicted to be potent by MultiCASE and the other two predictions from MultiCASE were not considered confident enough to make a prediction (Table 5). Ochratoxin A was one of the compounds that had a TD$_{50}$ value that was far more potent (956x) than predicted. Mycotoxins have complex structures that are difficult to accurately predict using *in silico* models. Cyanazine was also predicted to be less potent for reasons that are unclear, but the ratio of predicted TD$_{50}$ over experimental TD$_{50}$ (17x) was not as high the ratio for Ochratoxin A. Thus, while there may be some limitations for precisely predicting the pTD$_{50}$, the predictions are appropriately conservative, and an underestimation of risk is unlikely. Also, the use of multiple models can provide extra protection if one model underestimated risk (Matthews *et al.* 2008).

Decision Tree

Based on the results of the analysis, a proposed decision tree was developed (Figure 2). If the impurity is not determined to be genotoxic, then normal ICH Q3A / Q3B guidelines for impurities should be followed (ICH 2006a, b). However if genotoxicity is determined, then these models would be helpful for developing a safe limit. If the

prediction is potent in one model (either MultiCASE / VISDOM rat and mouse models) or no prediction can be made with either model, then the default TTC approach is applied. If the prediction is instead not potent in all models where confident predictions are made, then a numerical estimate is made using VISDOM. The highest calculated $p\text{TD}_{50}$ from either species is selected to determine the risk specific dose (RSD). The $p\text{TD}_{50}$ is converted back to the TD_{50}. The RSD is developed from the TD_{50} using linear-low dose extrapolation and assuming a 1 in 100,000 excess cancer risk. RSD is determined from the TD_{50} by Equation 2.

$$\text{RSD}\ (\mu\text{g/day}) = \left(\frac{\text{TD}_{50}\ (\text{mg/kg/day})\ \text{x}\ 0.00001}{0.5} \right) \text{x}\ 70\ \text{kg}\ \text{x}\ \frac{1000\ \mu\text{g}}{\text{mg}} \quad (2)$$

Case Study

Figure 3 represents a case study using the models developed in the study. Isopropyl chloride is an alkyl halide that is positive in the Ames assay. No carcinogenicity study has been published for isopropyl chloride to develop a risk estimate. The chronic limit using the default TTC approach is 1.5 µg/day. Both VISDOM and MultiCASE predicted that isopropyl chloride was not a potent carcinogen. The highest $p\text{TD}_{50}$ from both species was 3.47. This converts to a TD_{50} of 26.1 mg/kg/day resulting in an RSD of 37 µg/day. Assuming a 1 g drug substance/day dose, a limit of 37 ppm was developed. In comparison, the default TTC resulted in a limit of 1.5 ppm.

Control of short-term exposure to GTIs during clinical trials is a recent development in regulatory guidances, allowing for higher daily exposures during clinical trials (CHMP 2008A; USFDA 2008a). This has been coined as the “Staged-TTC”, which takes into consideration that lifetime exposure can be redistributed over shorter durations. The Staged-TTC uses more conservative assumptions based on the patient population which is a lower excess cancer risk level (1 in 1 million) and a 2-fold reduction in dose based on uncertainties from this linear redistribution of lifetime exposure. The

same assumptions for the Staged-TTC would still apply for the RSD during clinical trials. Table 6 shows a new staged-approach for isopropyl chloride using the new RSD compared to the defaulted Staged-TTC. The new staged-approach for isopropyl chloride takes into account similar assumptions as the Staged-TTC. The Staged-TTC doses following an *in silico* assessment are 25-times greater than the defaulted values, allowing limits for clinical trial manufacturing that more realistically represents the predicted potency of the compound. Although a high daily exposure would be considered acceptable, the ALARP principle and other regulatory guidelines (i.e. ICH Q3A or Q3B) would probably restrict exposures for short-duration use of clinical material (CHMP 2006; ICH 2006a, b).

Discussion

The resulting *in silico* model is an extension of the current EMEA and USFDA (draft) guidances in that it develops an effective prediction of a TD_{50}, which is an estimate of carcinogenic potency in lieu of defaulting to the conservative TTC (CHMP 2006; USFDA 2008a). The difficulty of extrapolating risk using the TTC is that it assumes an impurity with limited data is a potent carcinogen (Kroes *et al.* 2004). However, predicting carcinogenic potency can provide a better estimate of risk and allow for a more realistic estimate of acceptable exposure.

The USFDA draft document states "…the conduct of an SAR evaluation of an impurity may provide useful information. When a significant structural similarity to a known carcinogen is identified, the drug substance and drug product acceptance criteria (typically in units of parts per million or percent) can be set at a level that is commensurate with the risk assessment specific to that of the known compound (USFDA 2008a)." This approach is an expansion of the guidance by providing an *in silico* approach for estimating carcinogenic risk from a database of carcinogens. It is less dependent on selecting a "model carcinogen" because seemingly similar

carcinogens may have vastly different potencies. In the case of isopropyl chloride, there are a number of alkyl halides (e.g. ethyl chloride, 1,2-dichloropropane, etc.) that have carcinogenicity data and selecting the most similar compound based on expected biological effects would be difficult. There are also other scientific advantages that computerized models can provide such as machine learning, or taking into account the chemical reactivity of the molecule where visual inspection of the structure may miss these effects.

What is also important to highlight is that there are limitations to the model and it normally errors on the conservative side. This is indicated by the high sensitivity and low specificity in some cases. In other words, it is likely that a GTI will be flagged as potent, when it is not potent (false positive). Excluding the experimentally indeterminate range for VISDOM dramatically improved its sensitivity, so the decision tree considered indeterminate predictions as inadequate predictions. Regression analysis is useful for quantifying the TD_{50} of impurities predicted to be not potent. It is possible that a few dramatic missed predictions resulted in the low Q^2 for rats in the VISDOM model. The predictions were normally conservative, with the majority of predictions ≤ 5-fold from the actual carcinogenic potency. VISDOM predicted some of the compounds in the test set to be far less potent than experimentally derived results, but MultiCASE was normally able to correctly identify that those compounds were potent. However, Ochratoxin A was far more potent than predicted by VISDOM, and MultiCASE did not identify it as potent. Ochratoxin A is representative of a class of structurally complex compounds (mycotoxins) that is likely to result in misclassifications.

The case study and decision tree are intended as a guide to show how the *in silico* model can be used in practice for GTIs. The decision tree or process for applying the information may differ from organization to organization, but the principles of risk assessment and *in silico* modeling should stay fairly consistent.

The *in silico* analysis of isopropyl chloride, an Ames positive compound, predicted the compound to be of lower carcinogenic potency. This resulted in a higher acceptable daily intake when compared to the TTC for both chronic and short-term (e.g. clinical trial) exposures. However, while this risk assessment may be considered appropriate for carcinogenicity, there may be non-carcinogenic effects that should also be considered. The non-carcinogenic effect levels will be unknown for many GTIs, which is no different than many nongenotoxic impurities. Therefore, it is important that even though higher acceptable daily intakes may be allowed, the ICH Q3A / Q3B guidelines should be followed, which provide a framework for qualification of impurities (ICH 2006a, b).

We recognize that the use of *in silico* models to predict carcinogenic potency has applicability beyond the pharmaceutical GTIs. Benfenati *et al.* (2009) indicated the need for the prediction of carcinogenic potency in the food industry, especially for inadvertent or accidental presence of a chemical in products (Benfenati *et al.* 2009). As the application of the TTC continues to grow (e.g. food, pharmaceuticals, personal care products, etc.) so does the need to improve the accuracy of the risk assessment instead of assuming the worst-case (Blackburn *et al.* 2005; Humfrey 2007; Kroes and Kozianowski 2002; Kroes *et al.* 2004).

In conclusion, while *in silico* models have been effective for predicting threshold toxicity endpoints, this report shows the utility of *in silico* models in a step-wise approach for prediction of carcinogenic potency of GTIs as applied to pharmaceutical products. This approach can provide a more realistic estimate of cancer potency for GTIs rather than relying only on a single conservative estimation such as the TTC. While the approach has limitations, it was developed to be normally conservative to ensure drug safety and quality.

Tables

Table 1. Distribution of carcinogenicity data.

Species	Total Compounds (Training Set / Testing Set)	Potent (pTD50 ≥ 4.53)	
		Training	Testing
Rat	511 (460 / 51)	186 (40%)	25 (49%)
Mouse	402 (362 / 40)	74 (20%)	8 (20%)

Data retrieved from the Carcinogenic Potency Database (http://potency.berkeley.edu/).

Table 2. Validation statistics for the MultiCASE and VISDOM models. Indeterminate values were excluded and included from the test set.

Confident calls only, excluding indeterminate compounds						
	MultiCASE		VISDOM		Consensus	
	Rat	Mouse	Rat	Mouse	Rat	Mouse
Sensitivity	94%	86%	85%	60%	86%	88%
Specificity	50%	81%	44%	100%	36%	85%
Concordance	80%	83%	68%	90%	69%	86%
Applicability	49%	58%	43%	53%	63%	70%
PPV	80%	67%	69%	100%	72%	70%
NPV	80%	93%	67%	89%	57%	94%
All calls, including indeterminate compounds						
	MultiCASE		VISDOM		Consensus	
	Rat	Mouse	Rat	Mouse	Rat	Mouse
Sensitivity	69%	50%	81%	45%	71%	53%
Specificity	62%	74%	42%	100%	40%	75%
Concordance	67%	65%	69%	81%	62%	67%
Applicability	88%	93%	76%	78%	98%	98%
PPV	81%	54%	76%	100%	74%	57%
NPV	44%	71%	50%	77%	38%	72%

* Validation Parameters.

Sensitivity: The proportion of compounds correctly predicted to be potent relative to all compounds experimentally determined to be potent.

Specificity: The proportion of compounds correctly predicted to be not potent relative to all compounds experimentally determined not to be potent.

Concordance: The proportion of compounds correctly predicted to be potent and not potent relative to total number of predictions.

Applicability: Total predictions vs. total compounds tested.

Positive prediction value (PPV): The proportion of compounds correctly predicted to be potent relative to all predictions categorized as potent.

Negative prediction value (NPV): The proportion of compounds correctly predicted to be not potent relative to all predictions categorized as not potent.

Table 3. The model fit and resubstitution statistics for VISDOM.

Model	**R^2**	**R^2 adjusted**	**Q^2**
Mouse	0.655	0.618	0.615
Rat	0.790	0.750	0.335

R^2 - Correlation coefficient for training set and predictions of the training set $p\mathrm{TD}_{50}$s.

R^2 adjusted - Adjusts the R^2 for the number explanatory terms in the model.

Q^2 - Correlation coefficient for test set and predictions of the test set $p\mathrm{TD}_{50}$s.

Table 4. Percentage of compounds where the ratio was less than or equal to 1, 2, 5, and 10-fold from the experimental compound.

	VISDOM Percent of Compounds ≤Ratio (Predicted TD50 / Experimental TD50)			
Species	**≤ 1-fold**	**≤ 2-fold**	**≤ 5-fold**	**≤ 10-fold**
Rat	59%	64%	86%	86%
Mouse	66%	81%	88%	97%

Table 5. MultiCASE predictions for compounds where the VISDOM prediction of TD50 was greater than 10-fold higher than the experimental value.

Compound	Model	Ratio = $PredTD_{50}$ / $ExpTD_{50}$	MultiCASE Prediction
Cyanazine	Rat	17.5 = 110.4 mg/kg / 6.3 mg/kg	NA
Ochratoxin A	Rat	956 = 133.9 mg/kg / 0.14 mg/kg	NA
N-Nitrosoethylurethane	Rat	24= 2.14 mg/kg / 0.09 mg/kg	Potent
Chlorambucil	Mouse	586 = 76.2 m/kg / 0.13 mg/kg	Potent

Table 6. Daily Dose of Staged-TTC for Isopropyl Chloride Based on Defaulted Guidelines Versus *In Silico* Prediction.

Approach	Single Dose	≤ 1 month	≤ 3 months	≤ 6 months	≤ 12 months
Defaulted Staged-TTC (µg/day)[a]	120 µg	60 µg	20 µg	10 µg	5 µg
In Silico Staged-TTC (µg/day)[b,c]	3,151 µg	1,576 µg	525 µg	263 µg	131 µg

a. Defaulted Staged-TTC values based on EMEA CHMP Q&A (2008).

b. Lifetime exposure at a 1 in 1 million excess cancer risk is 94.5 mg/lifetime (3.7 µg/day x 365 days x 70 years) for the *in silico* prediction of isopropyl chloride. Calculations were modeled after the EMEA's defaulted Staged-TTC (with assumptions such as 2x reduction for uncertainties with linear adjustment of dose during short-term exposure). In comparison, lifetime exposure to the defaulted TTC is 3.8 mg/lifetime (0.15 µg/day x 365 days x 70 years).

c. ICH Q3A and Q3B limits must also be followed and would probably restrict actual exposure levels.

Figures

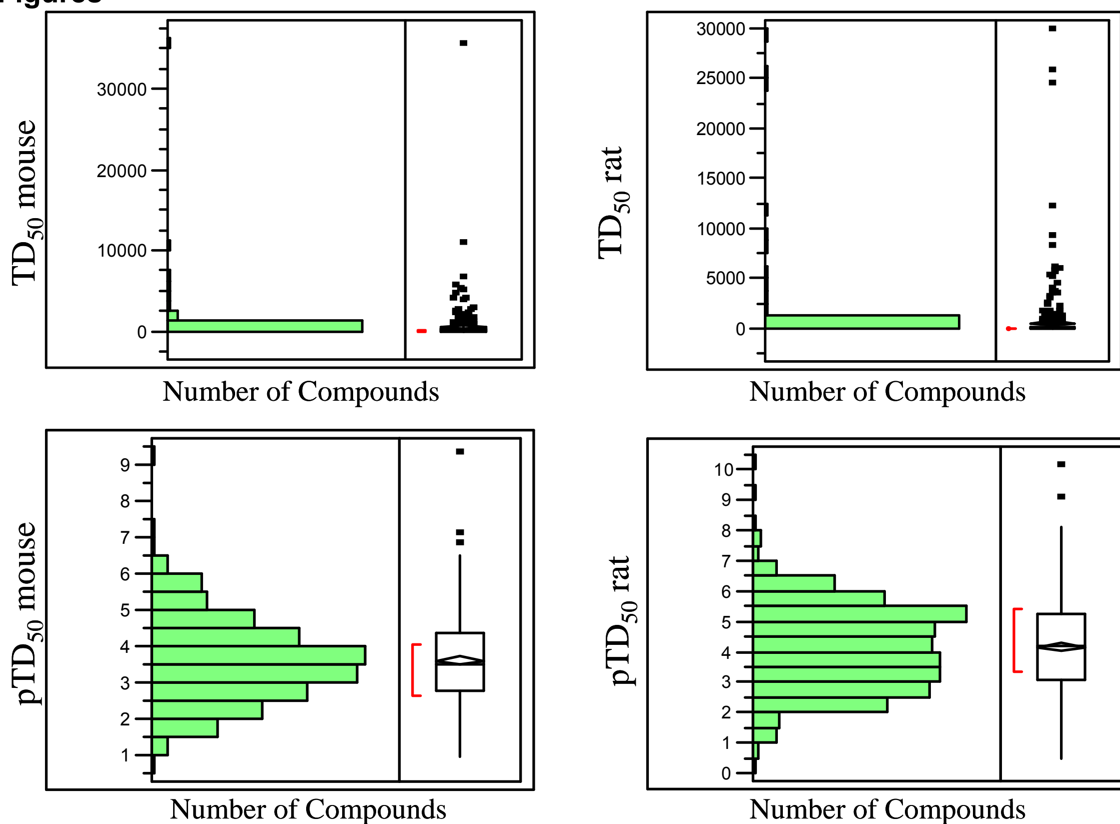


Figure 1. Distribution of the TD_{50} and pTD_{50} data for both rats and mice demonstrating the advantages of the data transformation.

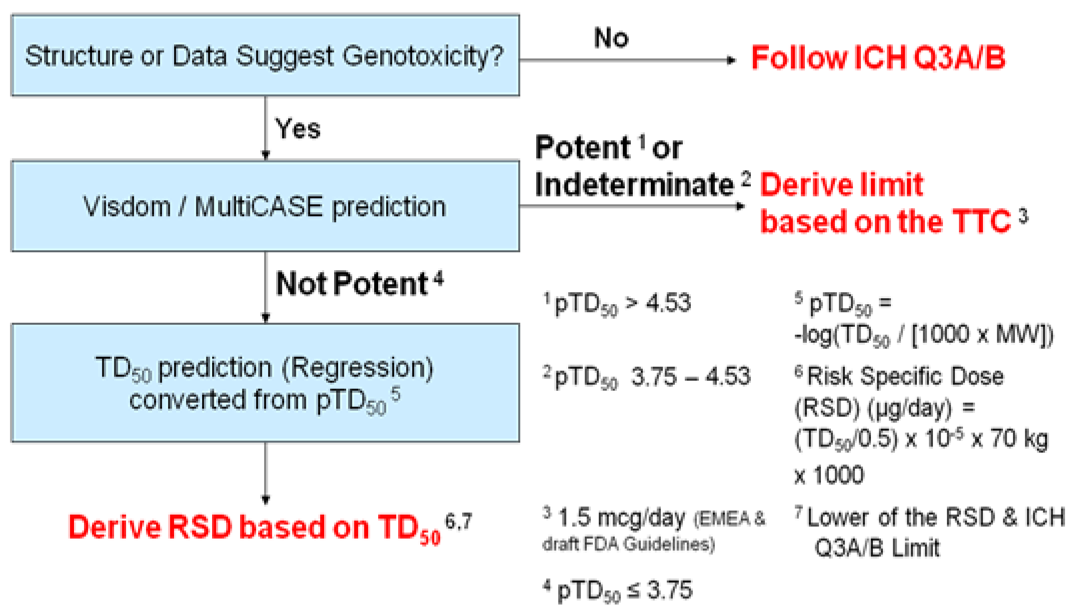
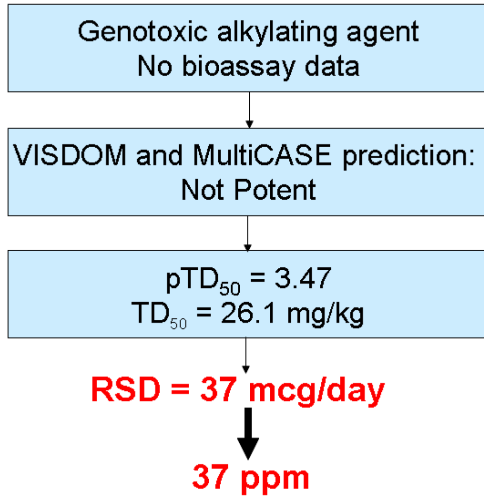


Figure 2. Proposed decision tree.

Isopropyl Chloride (CAS No. 75-29-6), SMILES: CC(C)Cl
Limits relative to a 1 g dose of drug substance:

Following Decision Tree for QSAR Approach:



Following Current Guidances

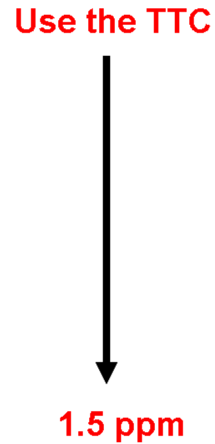


Figure 3. Case study using the models developed in the study.

QUANTITATIVE ASSESSMENT OF CUMULATIVE CARCINOGENIC RISK FOR MULTIPLE GENOTOXIC IMPURITIES IN A NEW DRUG SUBSTANCE[1]

Abstract

In pharmaceutical development, significant effort is made to minimize the carcinogenic potential of new drug substances (NDS). This involves appropriate genotoxicity and carcinogenicity testing of the NDS, and understanding the genotoxic potential of its impurities. Current available guidance recommends the use of the threshold of toxicological concern (TTC) for a single impurity where mutagenicity but no carcinogenicity information exists. Despite best efforts, the presence of more than one genotoxic impurity in an NDS may occur at trace levels. This paper repeats the analysis performed by others for a single genotoxic compound, but also uses statistical simulations to assess the impact on cancer risk for a mixture of genotoxic compounds. In summary, with the addition of multiple impurities all controlled to the TTC, an increase in cancer risk was observed. This increase is relatively small when considering the conservative assumptions of the TTC. If structurally similar compounds had an assumed strong correlation (±10 fold from the first randomly selected impurity) in cancer potency, the resulting cancer risk was not negatively impacted. Findings based on probabilistic analysis here can be very useful in making appropriate decisions about risk management of multiple genotoxic impurities measured in the final drug substance.

Introduction

New drug substances (NDSs) are extensively tested in toxicology studies before they enter the market. These tests can identify the risks associated with an NDS such as mutagenic, reproductive, nonneoplastic, and carcinogenic effects. A battery of genotoxicity tests is designed to detect damage to the DNA either through direct or

[1] This chapter has been published previously in: Bercu JP, Hoffman WP, Lee C, and Ness DK. (2008). Quantitative assessment of cumulative carcinogenic risk for multiple genotoxic impurities in a new drug substance. Regul. Toxicol. Pharmacol. 51, 270-7.

indirect mechanisms (ICH 1997a). If the compound is intended for chronic use, then a bioassay is performed on the compound to determine its carcinogenic potential (ICH 1997b).

Also tested along with the NDS are its impurities. Although these impurities may only exist at low levels, they are still evaluated for their potential toxicity (ICH 2006a). For impurities with unknown toxicity, nonclinical safety tests for the NDS can be used to qualify the impurities. It may be necessary to keep these impurities at lower levels for compounds that are considered “unusually toxic”.

It is assumed that the toxicity of genotoxic carcinogens has no threshold, requiring lower impurity limits than noncarcinogens (Barlow *et al.* 2006; USEPA 2005a). Tools such as *in silico* models, structure activity analysis, or genotoxicity tests can provide predictions of carcinogenicity for impurities without the cost or expense of a 2-year bioassay (Dobo *et al.* 2006; Hayashi *et al.* 2005; Matthews *et al.* 2006b). Efforts are made to remove genotoxic impurities from the synthesis of a molecule, but in some instances it is not technically feasible. If it is known that there is a potential for a genotoxic impurity, then the chemical process is designed to remove the impurity to a safe level. Furthermore, analytical chemistry confirms the process is efficient in removing genotoxic impurities (Argentine *et al.* 2007).

Safety evaluations are available for a genotoxic impurity with no carcinogenicity information. Since carcinogenic risk cannot be estimated with suitable precision or accuracy, the threshold of toxicological concern (TTC) has been applied to impurities where mutagenicity but not carcinogenicity is known about the impurity (Müller *et al.* 2006). This approach is similar to what has been applied to impurities in food (Kroes *et al.* 2000; Kroes *et al.* 2004; Munro *et al.* 1999). The TTC is the highest dose (µg/day) at which, with a high probability, the resulting increase in cancer risk over background is negligible (Müller *et al.* 2006). It was developed from analyzing the cancer potency

database for known carcinogens (Gold *et al.* 1999; Gold *et al.* 1991). From this analysis of carcinogens, certain classes of compounds were identified as being of higher carcinogenic potency (Kroes *et al.* 2004). These compounds, labeled as Cohorts of Concern (COCs), were predicted to have such high carcinogenic potency that their acceptable dose would be lower than the TTC. Therefore, the TTC is not recommended for COCs (CHMP 2006).

The current risk assessment approach is for a single genotoxic impurity. Multiple genotoxic impurities can increase the complexity of the risk analysis. There are no quantitative risk analyses for mixtures of genotoxic impurities, without carcinogenicity information, to determine their carcinogenic risks. However, for genotoxic carcinogens, USEPA has developed risk analyses for mixtures with exposure from hazardous waste sites. Response addition is recommended for carcinogens, where the mechanism of interaction among carcinogens is unknown (USEPA 2000b). A practical approach for the risk management of multiple genotoxic impurities can help ensure safe levels of impurities. The TTC relies on several conservative assumptions because of the uncertainties from estimating a cancer risk value when no oncogenicity study exists (Humfrey 2007; Munro *et al.* 1999). Development of a pharmaceutical takes several intermediate steps all of which can have reaction byproducts. It has been estimated that 20-25% of all intermediates are potentially genotoxic (Delaney 2007). However, most intermediates/byproducts are removed, and unlikely to carry through to the final drug substance. In instances where genotoxic impurities are measured in the final drug substance, an option is to incorporate a risk assessment approach for these impurities based on cumulative cancer risk.

This paper evaluated cancer risk of multiple genotoxic impurities by considering how the following factors impact the cancer risk: (1) co-administration of up to three genotoxic impurities, (2) the proportion of carcinogenic genotoxic impurities among all

genotoxic impurities, and (3) structural similarity among multiple genotoxic impurities. Finally, the potential utility of this analysis in determining acceptable daily intakes of multiple genotoxic impurities in an NDS was discussed.

Methods

Data source

The compounds used for this analysis were selected from the Gold carcinogenicity database (http://potency.berkeley.edu). This database includes the TD50s of those chemicals tested in a two-year carcinogenicity study. A total of 756 chemicals were considered carcinogenic in rats and/or mice.

In general, for a selected tumor at a target site, a portion of a population will develop this tumor spontaneously without exposure to any carcinogens. For the rest of the population, the proportion that eventually develops the tumor or remains tumor-free for a lifetime depends on the magnitude of exposure to a specific compound. For each compound and a selected tumor, the average daily dose at which 50% of the population will stay tumor free for a lifetime is defined as the TD50 (Peto *et al.* 1984). If a compound was declared carcinogenic from more than one experiment within the same species at the same target site, the resulting TD50 of the species and target was reported as the harmonic mean of the TD50s (http://potency.berkeley.edu/td50harmonicmean.html). For each species, the TD50 from the most potent target site was reported. If the compound was declared carcinogenic in both rats and mice, the lower TD50 was reported.

Construction of Cancer Potency Database (CPD)

The developed cancer potency database (CPD) contained 756 carcinogens from the Gold carcinogenicity database as a representative distribution of carcinogens. This approach is consistent with Kroes *et al.* (2004) and Fiori and Meyerhoff (2002), who used a similar dataset to establish an acceptable threshold for compounds which are

genotoxic or structurally alerting to be genotoxic. The intent of this paper was to use the same database used for a single impurity and extend its application to multiple impurities.

Since not all genotoxic compounds are carcinogenic, it was expanded to include additional entries which represented genotoxic noncarcinogens. Naturally, the cancer risk for a genotoxic noncarcinogen is 0. In this paper, to demonstrate how the probability that a genotoxic compound is a carcinogen impacts the cumulative cancer risk of multiple genotoxic impurities, 10, 50 and 80% of the compounds in the CPD were assumed to be carcinogenic. For example, a constructed CPD with 80% of the carcinogenic genotoxic compounds included 756 nonzero risks and 189 zero risks. Similarly, a constructed CPD with 50% of the carcinogenic genotoxic compounds included 756 nonzero risks and 756 zero risks.

The CHMP (2007) guidance on genotoxic impurities recommends that COCs be excluded from the use of the TTC. The identified COCs were aflatoxins, N-nitrosos, azoxys, steroids, and tetrahalogenated dibenzodioxins and dibenzofurans. The number of COCs in each structural alert category and the total in the CPD for the risk analysis are reasonably consistent with those in previous evaluations (Table 1) (Kroes *et al.* 2004). These COCs were removed from the database of carcinogens.

To determine the cumulative cancer risk of multiple impurities, it was important to understand if there is a relationship between structural alerts and cancer risks of compounds. Eighteen general categories of structural alerts have been established previously (Ashby and Tennant 1991; Cheeseman *et al.* 1999; Kroes *et al.* 2004). A compound could have multiple structural alerts and therefore could be in more than one structural alert category.

There may be an underlying mechanism such that certain structurally similar genotoxic impurities co-exist. The implication is that the cancer risks of those co-existing

genotoxic impurities would fall within a certain risk range. Therefore, an assumed risk range was selected which was 10-fold from the first impurity (100-fold total risk range). In comparison, the risk range for a randomly selected carcinogenic impurity is over 100-million fold.

Statistical Analysis

The primary purpose of this analysis was to establish the cumulative cancer risk profile of up to three genotoxic impurities. This was accomplished under each of the 6 scenarios resulting from the combination of the following two factors,

- three selected proportions of genotoxic impurities that are carcinogenic: 10, 50, and 80%,
- two different risk inclusion rules: one without restriction and the other with the 10-fold risk range.

Based on the TD50s of genotoxic impurities given in the database, the cancer risks were derived under some assumptions. Cumulative cancer risk profiles were then established for multiple genotoxic impurities using Monte Carlo simulations. The percentiles of interest of these profiles were obtained.

Derivation of the cancer risk from TD50 in mg/kg/day

Based on the given TD50s in the database and a general assumption of a linear relationship between the concentration (mg/kg/day) of the impurity and the cancer risk in Equation 1, the cancer risk value for a 70 kg subject receiving 1.5 µg/day of a genotoxic impurity was derived.

$$\frac{Risk1}{Concentration1} = \frac{Risk2}{Concentration2}. \qquad (1)$$

For an impurity, select the TD50 in mg/kg/day as Concentration 1, then Risk 1 is 0.5. Based on Equation 1, the cancer risk at Concentration 2 is

$$Risk2 = \left(\frac{Risk1}{Concentration1}\right) * Concentration2$$

$$= \left(\frac{0.5}{TD50(mg/kg/day)}\right) * Concentration2\ . \qquad (2)$$

Concentration 2 in ng/kg/day for a 70 kg subject receiving 1.5 μg of a genotoxic impurity per day is 21.43 ng/kg/day as calculated in Equation 3 below.

$$Concentration2 = \left(\frac{1.5\ \mu\text{g/day} * (1{,}000\text{ng}/\mu\text{g})}{70kg}\right). \qquad (3)$$

Substitute (3) into (2), the corresponding cancer risk, Risk 2, is

$$Risk2 = \left(\frac{0.5}{TD50(mg/kg/day) * (1{,}000{,}000ng/mg)}\right) * \left(\frac{1.5\ \mu g/day * (1{,}000ng/\mu g)}{70kg}\right).$$

Or for a given genotoxic impurity with a TD50 in mg/kg/day, the cancer risk for a 70 kg subject receiving 1.5 μg/day of it is

$$Risk2 = \frac{1.5}{140{,}000 * TD50}\ . \qquad (4)$$

Calculation of cumulative cancer risks associated with multiple genotoxic impurities

A special relationship between the administration of dose d_A of Impurity A and the corresponding probability p_A of developing cancer is the cumulative distribution function (CDF) (Hogg *et al.* 2005). Similarly, for Impurity B, there is a CDF that links the cancer risk, p_B, to the dose d_B of Impurity B. For the joint administration of Impurity A at dose d_A and Impurity B at dose d_B, there is a joint CDF of these two impurities that gives the total cancer risk. However, knowing the individual CDFs is not enough to uniquely define the joint CDF. For example, if both individual CDFs are normal distributions, it would still require the correlation coefficient between the two to determine the joint CDF. At this point, there are no known joint CDFs for the cancer risks of these impurities in the literature. This is likely due to the unknown relationship among the cancer risks of the impurities and extensive resources needed to conduct proper studies to estimate joint cancer risks.

If the true joint CDF of the cancer risk of two impurities, A and B, is a surface that slowly rises at low doses and climbs up to eventually approach 1 as doses increase, then one can approximate it with a plane in the shallow rising region, the region of very low doses of Impurity A and Impurity B. In a three-dimensional space made up of (x, y, z)'s, a general equation for a plane that passes through the origin can be expressed by

$$\frac{x}{a} + \frac{y}{b} + \frac{z}{c} = 0, \text{ where a, b and c are nonzero constants.} \quad (5)$$

To approximate the slow rising region of the joint cancer risk of Impurities A and B, we can further restrict constants a, b and c to satisfy $ab>0$ and $ac<0$ to avoid negative cancer risks.

Given that the approximated cancer risk at dose d_A of Impurity A is p_A and at dose d_B of Impurity B is p_B, both $(d_A, 0, p_A)$ and $(0, d_B, p_B)$ should satisfy Equation 5 as

$$\frac{d_A}{a} + \frac{0}{b} + \frac{p_A}{c} = 0, \text{ and} \qquad (6)$$

$$\frac{0}{a} + \frac{d_B}{b} + \frac{p_B}{c} = 0. \qquad (7)$$

Summing up Equations 6 and 7 leads to the following,

$$\frac{d_A}{a} + \frac{d_B}{b} + \frac{p_A + p_B}{c} = 0. \qquad (8)$$

Since (d_A, d_B, $p_A + p_B$) satisfies Equation 5, the joint cancer risk of these two impurities is the sum of the individual cancer risks, $p_A + p_B$. Based on this rationale, the joint (or cumulative) cancer risk of multiple impurities in the low risk region can be approximated by adding the individual cancer risks. This is consistent with the response addition used for carcinogens with unknown interactions (USEPA 2000b).

Percentiles of cancer risks

The percentiles of cancer risks for one impurity at the TTC of 1.5 μg/day have been published in the literature (Munro *et al.* 1999). To compare the results from impurities in the CPD, the cancer risk corresponding to the same TTC of 1.5 μg/day was either 0 or calculated using Equation 4. Percentiles were estimated from the calculated cancer risks. For example, the 90th percentile is the value below which at most 90 percent of all the cancer risks fall and above which the remaining 10% can be found. For multiple genotoxic impurities each at 1.5 μg/day, the cumulative cancer risks were obtained by Monte Carlo simulations as follows.

Given a positive integer K with K > 1,

1. Randomly select one cancer risk from the set of cancer risks of all genotoxic impurities in the CPD.

2. Repeat the previous step K times.

3. Obtain the cumulative cancer risk for the combination of K impurities by summing the individual cancer risks.

4. Repeat all steps 20,000 times.

The 90th, 93rd and 95th percentiles were estimated from these 20,000 simulated cumulative cancer risks as described for one impurity. This exercise was carried out for the CPD with 10, 50 or 80% of carcinogenic genotoxic impurities.

Structural similarity in genotoxic impurities

Certain genotoxic impurities may have a higher chance to appear together due to the similarity in chemical structures. The cancer risks of these structurally similar impurities may be relatively closer than those that are not. Under these conjectures, the cumulative cancer risks for the combination of K structurally similar genotoxic impurities, K > 1, were obtained by first selecting a genotoxic impurity. Then allow only those with cancer risks within a 10-fold range of the selected one to be included in the set of K impurities and sum up these K cancer risks. Cumulative cancer risks for K structurally similar impurities were simulated by Monte Carlo simulations. If the first randomly selected genotoxic impurity was noncarcinogenic, hence a 0 cancer risk, then all structurally similar impurities would also have 0 cancer risks. Therefore, these noncarcinogenic genotoxic impurities were not randomly selected in the simulation; instead, cumulative risks of 0 were added back to the final set of 20,000 simulated cancer risks.

For multiple structurally similar genotoxic impurities each at 1.5 μg/day, the cumulative cancer risks were obtained by Monte Carlo simulations as follows.

Given a positive integer K with K > 1,

1. Randomly select one cancer risk from the subset of nonzero cancer risks of the CPD, say risk R.

2. Randomly select one nonzero cancer risk from the set that are within the closed interval of [R/10, 10R].

3. Repeat the previous step K-1 times.

4. Obtain the cumulative cancer risk for the combination of K impurities by summing the individual cancer risks.

5. Repeat all steps 20,000 times.

The 90th, 93rd and 95th percentiles were estimated from these 20,000 simulated cumulative cancer risks and all the 0 cancer risks as described for one impurity. This exercise was carried out for the CPD with 10, 50 or 80% of carcinogenic genotoxic impurities.

To better understand the association between the chemical structure and the cancer risk, a structure based categorization of the 756 carcinogenic compounds from the CPD was created. This process was aided by a database of structural characterizations for carcinogens provided by Dr. Kirk Arvidson (United States Food and Drug Administration). Each genotoxic compound was categorized as either unlabeled (no structural alerts) or some of the 18 structural alert groups. The cancer risks of the genotoxic impurities in each structural alert group were plotted in Figure 4 and ordered by their medians. The 2.5^{th} and 97.5^{th} percentiles of each of these 19 groups (18 structural alert groups and one unlabeled) were also plotted to assess the impact of the chemical structure on the cancer risk.

Results

Cancer risk of a single genotoxic impurity compared to literature

Our results for a single genotoxic impurity (Table 2) are consistent with previous analyses (Fiori and Meyerhoff 2002; Kroes *et al.* 2004; Munro *et al.* 1999). A survey of different analyses used to develop a risk specific dose for genotoxic compounds determined that the results are slightly but not significantly different from each other

(Fiori and Meyerhoff 2002). Our results show that the probability of not exceeding a 1 in 100,000 (10^{-5}) excess risk of cancer when exposure is at the TTC of 1.5 μg/day is 95%, assuming 50% of genotoxic chemicals are also carcinogenic. Munro *et al.* (1999) determined that the probability was 93% for the same dose, excess cancer risk, and assumptions but the database contained COCs. Kroes *et al.* (2004) also observed an increased probability when COCs were removed.

Cancer risk of multiple genotoxic impurities

As the number of genotoxic impurities increased, so did the excess cancer risk estimates (Table 2). The cumulative probabilities for cumulative cancer risks of up to three genotoxic impurities are plotted in Figure 1, assuming 50% of the impurities are carcinogenic and each impurity is controlled to 1.5 μg/day. The addition of a genotoxic impurity increased the 90th percentiles (Figure 1).

Impact of the assumption of probability that a genotoxic compound is carcinogenic

The probability that a genotoxic compound is also a carcinogen (predictive probability) had an impact on excess cancer risk estimation of a single genotoxic impurity (Row 1 in Table 3). This probability was set at 50% for genotoxic chemicals in risk assessment involving the TTC (Barlow *et al.* 2001; Müller *et al.* 2006). Higher or lower probabilities may be used with proper justifications. As this probability increased, the corresponding cancer risk also increased. For example, consider the CPD for a single impurity, as this probability increased from 10, 50 to 80%, the cancer risk at the TTC increased from 1.79×10^{-10}, 3.27×10^{-6}, to 7.71×10^{-6} accordingly. For multiple genotoxic impurities (Table 3), this probability has a similar effect on the excess cancer risk as it has on a single impurity.

Impact of related potencies for multiple genotoxic impurities

Structurally similar genotoxic impurities may have similar excess cancer risks after exposure at the TTC. For each category of alerting structures, the range of 2.5th to

97.5^{th} percentiles for excess cancer risk varied between 3 log orders up to 8 log orders except for aflatoxin (2.17 log) and alpha nitrofuryl compounds (2.59 log). The overlapping of the (2.5^{th}, 97.5^{th}) percentile intervals of the structural alert categories indicates a weak relationship between the structure and cancer risk (Figure 4). However, stronger relationships could exist with compounds that differ only by a few elements. Assuming a stronger relationship between the cancer risk and similar chemical structure, the cumulative cancer risks for multiple impurities are presented in Table 4 and plotted in Figure 2. This stronger relationship would only allow additions to a single impurity by those with cancer risks in the 10-fold range of the first impurity. This resulted in lower excess cancer risks than before. For example, if one assumes 50% of genotoxins are human carcinogens, then the 90th percentile of the cumulative cancer risk for three impurities at the TTC would decrease from $1.73x10^{-5}$ to $1.20x10^{-5}$. Recall that the probability that a genotoxic compound is also a carcinogen (predictive probability) has an impact on excess cancer risk estimation of a single genotoxic impurity (Row 1 in Table 3 or Table 5). As the predictive probability increased from 10, 50 to 80%, the cumulative cancer risk of these impurities at the TTC increased as well (Figures 3A-3C) for structurally similar impurities.

Discussion

If a genotoxic chemical has the potential to enter the synthesis of a drug substance, it is removed by process chemistry and analytical tests confirm removal (Argentine *et al.* 2007). Removal of the genotoxic compound must at a minimum ensure that the potential impurity is below a safe level.

Threshold genotoxic mechanisms may exist such as topoisomerase inhibition (Anderson 2004; Lynch *et al.* 2003), oxidative stress (Beddowes *et al.* 2003), aneuploidy from interaction with cell division (Arni and Hertner 1997; Decordier *et al.* 2002), and DNA synthesis inhibition (Galloway *et al.* 1998). Furthermore, *in vitro* studies with

methyl methanesulfonate and ethyl methanesulfonate indicate that a threshold may exist for some alkylating agents (Doak *et al.* 2007; Jenkins *et al.* 2005). If a genotoxic impurity has a threshold mechanism, then a permissible daily exposure (PDE) can be derived by dividing the no-observed effect level or lowest-observed effect level by uncertainty factors (CHMP 2006). When no threshold can be demonstrated, the TTC approach is applied to the genotoxic impurity.

Currently, the risk assessment framework surrounding multiple genotoxic compounds is limited. This can lead to concerns about how to control for these impurities in an NDS (Jacobson-Kram and Jacobs 2005). The quantitative risk analysis described in this report will help inform risk managers about acceptable levels for a mixture of genotoxic impurities.

We recognize that cancer risks reported in this paper may be overestimated. Establishment of the TTC is based on several conservative assumptions, which have been detailed in previous publications (Delaney 2007; Humfrey 2007; Krewski *et al.* 1990; Munro *et al.* 1999). These assumptions were applied so that a safe dose can be established from little toxicological information. Therefore, the slight increase in cancer risk observed in this report from multiple impurities seems relatively insignificant when considering the assumptions used in development of the TTC.

From our analysis, aflatoxins, azoxy compounds, steroids, N-nitroso compounds and halogenated diobenzo-p-dioxins and dibenzofurans were the most potent carcinogenic structural classes (Figure 4). This is consistent with Kroes *et al.* (2004) who labeled these compounds as COCs. The COCs and nonCOCs had variation surrounding cancer potency. While many of the COCs were potent carcinogens, some of the COCs were not. Furthermore, some compounds were highly potent even though they were not labeled as a COC. CHMP (2007) and Kroes *et al.* (2004) exclude

compounds in the COC structural class from the TTC risk analysis. COCs were excluded from our analysis to follow this guidance.

These cancer risks were estimated based on different assumptions about a genotoxic compound being carcinogenic. Genotoxicity tests, although useful tools to help identify potential carcinogens, are not 100% predictive. As the predictive probability for carcinogenicity of a genotoxicity test increases from 10 to 80%, the associated cancer risk also increases (Table 3), which is consistent with previous findings (Munro *et al.* 1999). Risk values at the 50th and 80th percentiles are considered the most relevant for risk assessment as they cover the range of predictability for a genotoxicity assay (Barlow *et al.* 2001; Kirkland *et al.* 2005; Matthews *et al.* 2006a).

In some instances a genotoxic impurity may be structurally similar to another impurity. The synthesis of an NDS is the reaction of compounds to form a pharmaceutical product. If a genotoxic compound is introduced in the synthesis of an NDS, this compound could become an impurity. In addition, other structurally similar reaction byproducts or subsequent compound intermediates could become impurities as well. As shown in Figure 4, for compounds with the same alerting structure, carcinogenic potency varied amongst compounds. Many of the compounds with the same alerting structure still differed substantially with the rest of the moiety, or had an additional structural alert. When considering the cumulative cancer risk of multiple, structurally-related impurities, the likelihood of toxicological synergy is unknown but is considered to be unlikely at these vanishingly small doses. Structurally similar compounds may have a similar mechanism of action, which may imply risk values that are close to each other. However, slight changes in structure could influence the ability for a compound to be absorbed, metabolized and penetrate into the target organ. The analysis compared the cumulative risk based on random potencies of multiple impurities with potencies that all fell within a $\pm$ 10-fold of first impurity. When cancer risk of multiple

genotoxic impurities was dependent on the first impurity, total cancer risk was decreased relative to total cancer risk for structurally-unrelated impurities. This resulted in a more favorable outcome. Therefore, even if it were assumed that genotoxic impurities related in structure had similar carcinogenic potencies, it should not result in lower safety limits than for compounds structurally unrelated to each other.

This mixture assessment assumes chronic exposure of the genotoxic impurity. It would not affect the assumptions used to derive the staged TTC approach, allowing higher daily exposures over shorter durations (Müller *et al.* 2006). As for single impurities, higher risk values and thus higher exposures may be considered for life-saving medicines such as those for the treatment of cancer or Alzheimer's disease (CHMP 2006). Finally, lower exposure to genotoxic impurities may be recommended for medicines that are indicated for children (USEPA 2005b).

Our risk assessment approach for multiple genotoxic impurities is consistent with risk assessment practices for mixtures of carcinogens, where response addition is currently applied (Kaldor and L'Abbe 1990; Kodell and Chen 1994; Krewski and Thomas 1992; USEPA 2000b). Analyses for mixtures of carcinogens have been limited to specific instances such as Superfund contamination (USEPA 2000b). The same principles were applied to genotoxic impurities, where the patient may be exposed to a mixture of genotoxic impurities along with the NDS. Practical decisions for the risk management of genotoxic impurity mixtures should consider the following important factors: low level of potential exposure, conservative nature of the TTC based on limited genotoxicity information, extensive safety testing of the drug substance containing the impurities, and the benefit a patient receives from consuming the pharmaceutical (Müller *et al.* 2006; Munro *et al.* 1999).

Conclusions

The formation of genotoxic impurities in the synthesis of an NDS should be avoided whether as a single impurity or multiple impurities. If removal of genotoxic impurities is not technically feasible, then a safety limit must be determined for a mixture of potential genotoxic impurities. The TTC is a generally accepted exposure for a single genotoxic impurity (CHMP 2006; Kroes *et al.* 2004; McGovern and Jacobson-Kram 2006; Müller *et al.* 2006). This paper furthers the discussion by using a probabilistic analysis to characterize cancer risk for a mixture of genotoxic impurities.

With the addition of one to two genotoxic impurities, a slight increase in cancer risk was observed. However, this minimal increase is not of concern given the conservative assumptions incorporated in the TTC. When considering structurally-related impurities, the cumulative cancer risk assessment is more favorable compared to unrelated impurities. This is based on the results of statistical simulation of compounds with related potencies and the expectation that toxicological synergy is not likely at these extremely low doses. Determining the acceptable number of genotoxic impurities warrants a broader discussion; however, this analysis suggests that up to three genotoxic impurities, whether structurally related or not, should be acceptable in most cases in pharmaceutical development. Four or more genotoxic impurities is a less likely scenario and should be discussed on a case-by-case basis.

Tables

Table 1. Cohorts of concern identified in current database.

Structure	Our Database	Kroes *et al.* 2004
Aflatoxin	5	5
N-nitroso	104	105
Azoxy	6	5
Steroid	13	11
tetrahalogenated dibenzodioxins and dibenzofurans	5	5
Total	133	131

Table 2. The percentiles of cumulative cancer risk of multiple genotoxic impurities assuming 50% of the genotoxic impurities are carcinogens.

No. of Impurities	Percentile of Cancer Risk		
	90%	93%	95%
1	3.27E-06	6.57E-06	1.03E-05
2	9.85E-06	1.50E-05	2.32E-05
3	1.73E-05	2.73E-05	3.79E-05

Note 1. Each impurity is controlled to a TTC of 1.5 µg/day.

Table 3. The 90th percentile of the cumulative cancer risks of multiple genotoxic impurities each with 10, 50 and 80% as carcinogens.

No. of Impurities	Percent of Carcinogens		
	10%	50%	80%
1	1.79E-10	3.27E-06	7.71E-06
2	2.67E-07	9.85E-06	1.79E-05
3	1.09E-06	1.73E-05	3.08E-05

Note 1. Each impurity is controlled to a TTC of 1.5 µg/day.

Note 2. For one impurity from a sample of 10% being carcinogens, i.e. 90% zero cancer risks and 10% nonzero risks, the 90th percentile is 1.79×10^{-10}, the midpoint between 0 and the lowest nonzero risk of 3.58×10^{-10}.

Table 4. The percentiles of cumulative cancer risks of multiple impurities with individual cancer risks contained within a 10-fold range. Assume 50% of the genotoxic impurities are carcinogens.

No. of Impurities	Percentile of Cancer Risk		
	90%	93%	95%
1	3.27E-06	6.57E-06	1.03E-05
2	7.70E-06	1.25E-05	1.92E-05
3	1.20E-05	1.83E-05	2.87E-05

Note 1. Each impurity is controlled to a TTC of 1.5 µg/day.

Table 5. The 90th percentile of the cumulative cancer risks of multiple genotoxic impurities each with 10, 50 and 80% as carcinogens and with individual cancer risks contained within a 10-fold range.

No. of Impurities	Percent of Carcinogens		
	10%	50%	80%
1	1.79E-10	3.27E-06	7.71E-06
2	9.83E-10	7.70E-06	1.46E-05
3	1.84E-09	1.20E-05	2.16E-05

Note 1. Each impurity is controlled to a TTC of 1.5 µg/day.

Note 2. For one impurity from a sample of 10% being carcinogens, i.e. 90% zero cancer risks and 10% nonzero risks, the 90th percentile is 1.79×10^{-10}, the midpoint between 0 and the lowest nonzero risk of 3.58×10^{-10}.

Figures

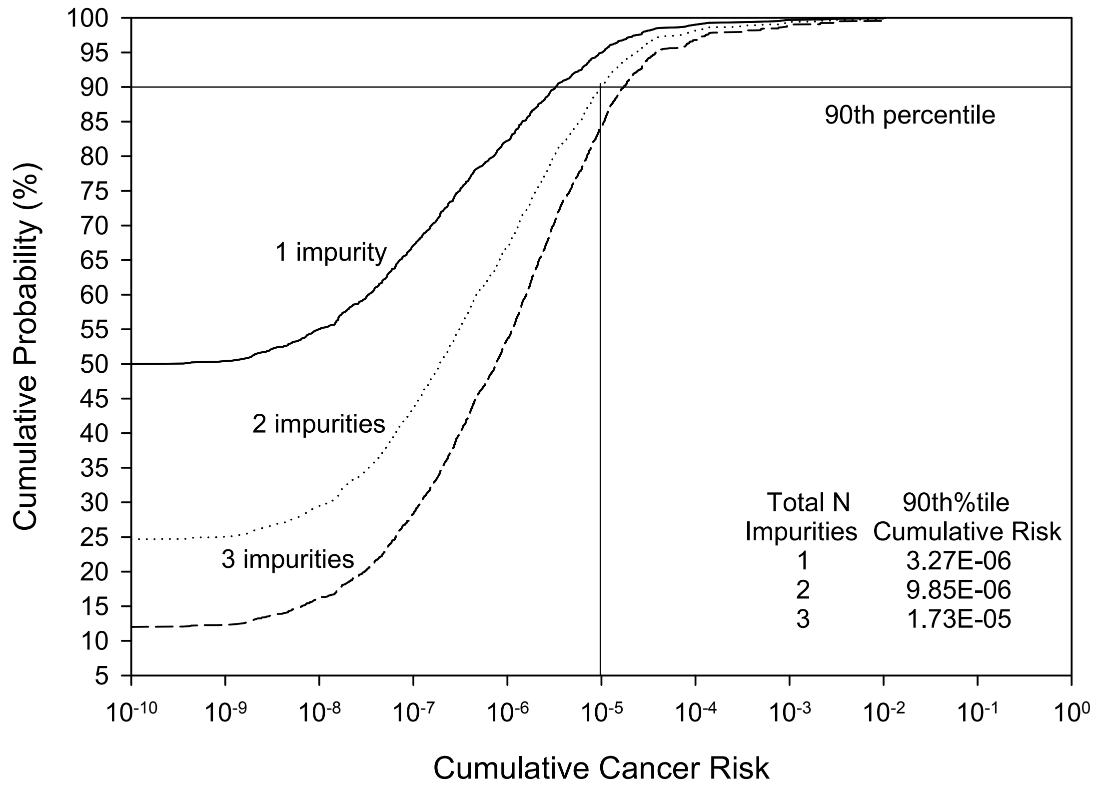


Figure 1. Cumulative probability of cumulative cancer risk for up to three genotoxic impurities. Assume 50% of the impurities are carcinogenic and each impurity is controlled to a TTC of 1.5 μg/day.

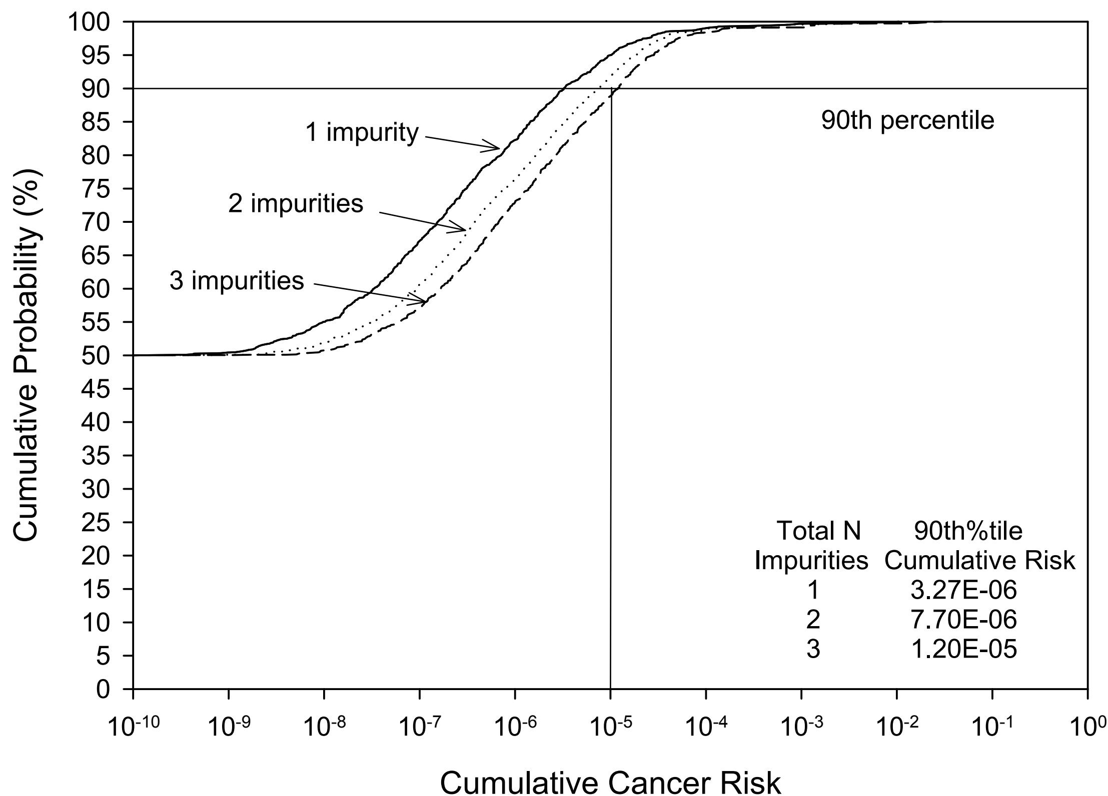
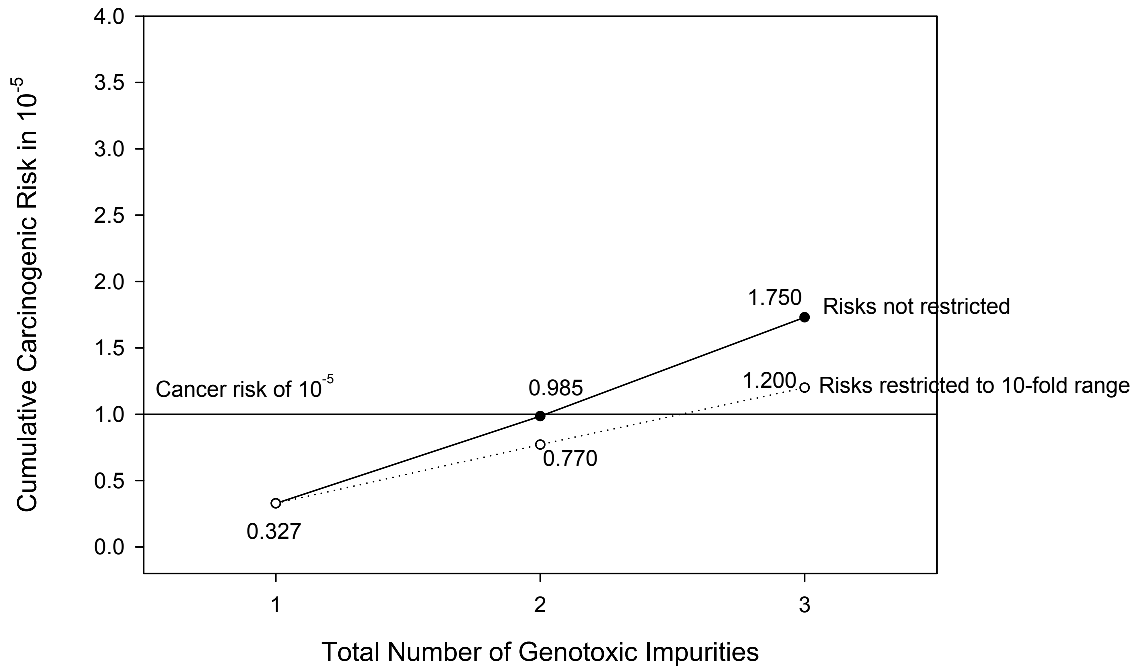
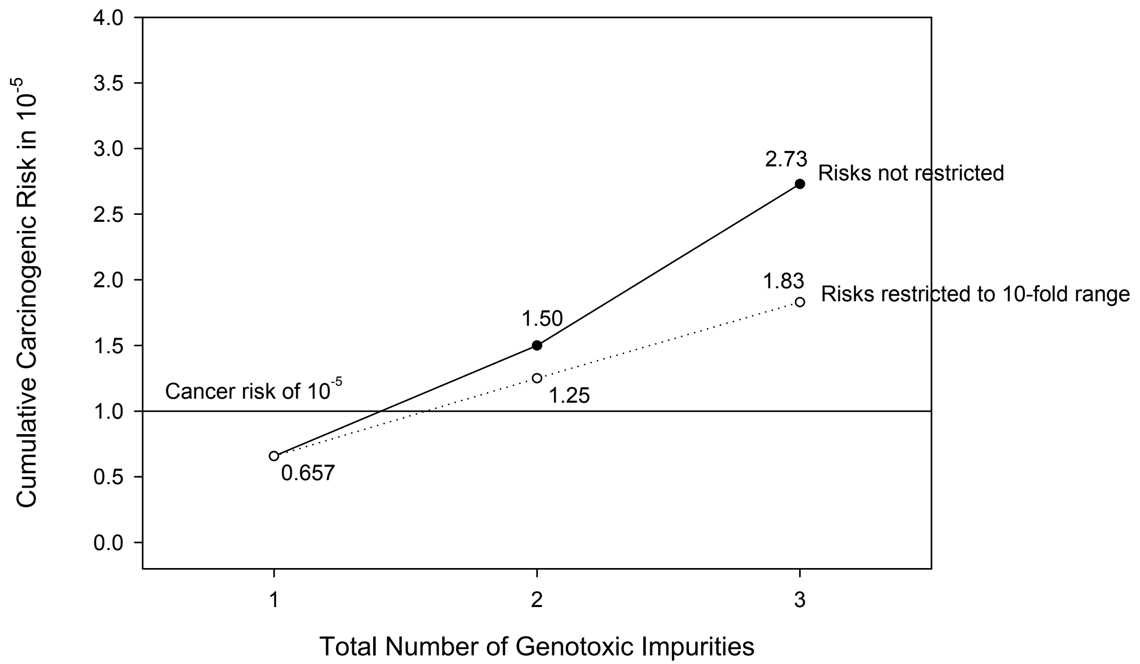


Figure 2. Cumulative probability of cumulative cancer risk for up to three genotoxic impurities with individual cancer risks contained within a 10-fold range. Assume 50% of the impurities are carcinogenic and each impurity is controlled to a TTC of 1.5 µg/day.

A.



B.



C.

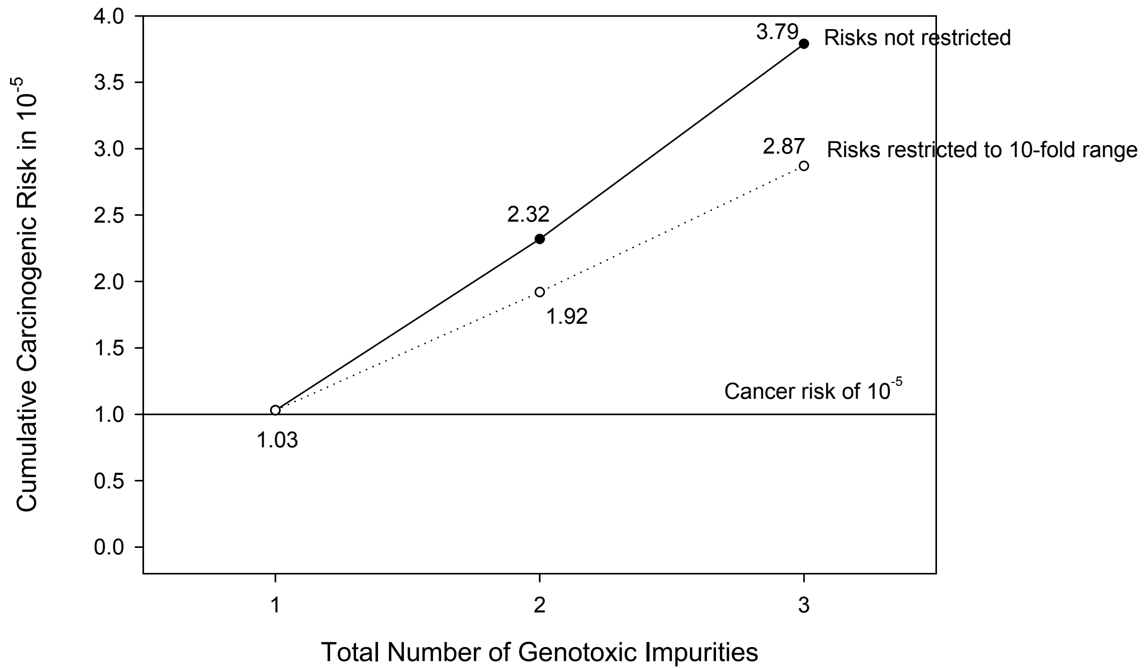


Figure 3. Cumulative cancer risk (10^{-5}) at the 90th(3A), 93rd (3B) and 95th (3C) percentile for up to three genotoxic impurities. Assume 50% of the impurities are carcinogenic and each impurity is controlled to a TTC of 1.5 µg/day.

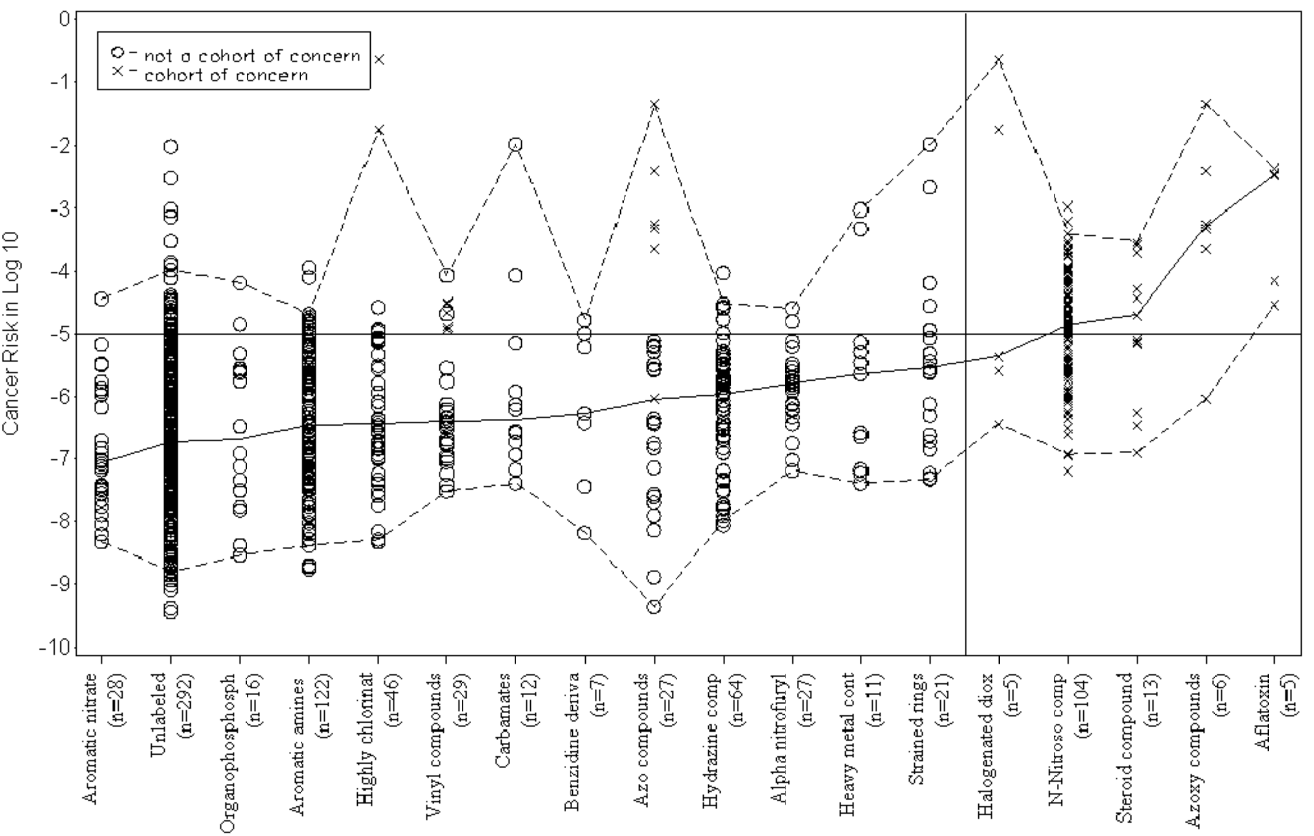


Figure 4. Cancer risk in Log10 with upper 97.5% and lower 2.5% limits (ordered by median) for each structural alert category. Exposure to each compound is set to a TTC of 1.5 µg/day.

Note: Some of the 756 impurities have multiple structural alerts; therefore, 835 cancer risks were plotted.

A STRATEGY FOR RISK ASSESSMENT OF HUMAN GENOTOXIC METABOLITES[1]

Abstract

The role of metabolism in genotoxicity and carcinogenicity of many chemicals is well established. Accordingly, the use of metabolic activation systems in genotoxic hazard testing is routinely applied when testing pharmaceutical candidates prior to clinical investigations. However, it is also well known that significant differences can exist between human metabolism and that which occurs with *in vitro* and *in vivo* genotoxicity tests. This poses challenges when considering the adequacy of hazard identification and cancer risk assessment if a human metabolite of genotoxic concern is identified during the course of drug development. Since such challenges are particularly problematic when realized in the later stages of drug development, a framework for conducting a carcinogenic risk assessment for human genotoxic metabolites is desirable. Here we propose a risk assessment method that is dependent upon the availability of quantitative human and rodent ADME (absorption, distribution, metabolism, excretion) data, such that absolute exposures to a metabolite of genotoxic concern can be estimated at the intended human efficacious dose and the maximum dose used in the 2-year rodent bioassay(s). The exposures are then applied to the risk assessment framework that allows one to understand the probability of a known or suspect genotoxic metabolite posing a carcinogenic risk in excess of 1 in 100,000. Practical case examples are presented to both illustrate the application of the risk assessment method within the context of drug development and to highlight its utility and limitations.

[1] This chapter has been published previously in: Dobo KL, Obach RS, Luffer-Atlas D, and Bercu JP. (2009). A Strategy for Risk Assessment for Human Genotoxic Metabolites. Chem. Res. Toxicol. 22, 348-56. Use of the manuscript was with permission from the primary author. Contributions to the manuscript were as follows: (1) developing a cancer risk assessment model for genotoxic metabolites, (2) mathematical evaluation of the case studies, (3) overall risk assessment strategy, and (4) supporting background material.

Introduction

In general, the non-clinical safety testing of a pharmaceutical drug candidate assesses four major toxic endpoints: (1) genotoxicity, (2) acute and/or chronic general target organ toxicity (3), acute and/or chronic reproductive toxicity, and lastly, (4) carcinogenicity. Since these toxicities are complex and often inter-dependent, there can be additional challenges that go beyond traditional non-clinical safety assessment for the parent drug. Recently there have been efforts to develop a common framework to manage the safety of drug metabolites, such that pharmaceutical sponsors and regulators have confidence that non-clinical safety evaluations conducted in support of clinical investigations adequately assess the safety of the full profile of metabolites to which humans are exposed. In the development of a framework for assessing metabolite safety, genetic safety warrants special consideration given that the role of metabolism in genotoxicity and carcinogenicity of chemicals is well established and that some mechanisms of genotoxicity and carcinogenicity are believed to occur through non-threshold mediated mechanisms of action. Furthermore, the ability of both *in vitro* and *in vivo* test systems to model human metabolism can be variable and therefore may limit the ability to adequately conduct human genotoxicity and carcinogenicity risk assessment in certain cases (Ku *et al.* 2007). Despite the need for evaluating genetic safety of metabolites, no practical recommendations are available to address human genotoxic metabolites that are identified during clinical development.

Some safety guidances have been developed with triggers for concern based on relative abundance of metabolites (Baillie *et al.* 2002; USFDA 2008b). Others have proposed a strategy in which absolute exposure to metabolites in humans triggers further consideration of metabolite safety (Smith and Obach 2005). A working group convened at the 4th International Workshop of Genotoxicity Tests (IWGT) acknowledged the need for a practical strategy to respond to documented human metabolite exposures

and suggested that an absolute exposure, a threshold of toxicological concern (TTC), be defined in order to better support a risk assessment approach for genotoxic human metabolites (Ku *et al.* 2007).

In traditional risk assessment, drug exposure is quantified so that its risk may be characterized in relation to a dose-response assessment. While relative abundance of metabolites can be used as a trigger for an assessment, absolute abundance enables exposure to be quantified to enable risk assessment. Excreted drug-related material can also be used to quantify the systemic body burden of a genotoxic substance, allowing for an evaluation of risk (Smith and Obach 2005, 2006). The most effective risk assessment should ideally encompass exposure to both circulating and excreted drug-related material.

Risk-based safety assessments have been applied to manage human exposure to genotoxic/carcinogenic compounds for numerous situations. Some of the basic principles include the assumption that compounds that are carcinogenic via a threshold mode of action have no adverse effects at low doses (Butterworth and Bogdanffy 1999; Butterworth *et al.* 1995; Clewell 2005; USEPA 2005a). In such cases, a no-adverse effect level (NOAEL) or lowest-adverse effect level (LOAEL) is identified from all relevant species tested to determine the point of departure. The point of departure is then divided by appropriate safety factors to derive the acceptable dose at which no deleterious effects are expected (Dourson 1996; Dourson and Stara 1983; USEPA 2002). In contrast, for genotoxic carcinogens, a non-threshold mode of action has been assumed which means there is a risk of cancer even from exposure to low doses (USEPA 2005a). As such, carcinogens in food, water, air, pharmaceuticals, etc. are limited so that the risk of cancer over background from its exposure is negligible (Barlow *et al.* 2006; ICH 1997c; Kroes *et al.* 2004; USEPA 1991). Typically, excess cancer risk is calculated by using very conservative, low-dose linear extrapolation techniques.

The most recent example of broad application of risk-based safety assessment within the pharmaceutical industry, is establishing allowable limits of exposure to genotoxic impurities present in active pharmaceutical ingredients. In general, impurities with genotoxic potential are considered "unusually toxic" and are not qualified by the thresholds used for other impurities (CHMP 2006; ICH 2006a; Jacobson-Kram and McGovern 2007; McGovern and Jacobson-Kram 2006). Instead, a risk-based approach has been applied to control genotoxic impurities to negligible levels (Fiori and Meyerhoff 2002; Kroes *et al.* 2004; Müller *et al.* 2006). Since carcinogenicity information is often not available for genotoxic impurities, an acceptable threshold (i.e. TTC) was determined by assessing the carcinogenic potencies from a large database of known carcinogens and setting an exposure limit to a dose that has a low likelihood of exceeding a negligible increase in cancer risk (CHMP 2006). This approach was conservative because it not only assumed low-dose linear extrapolation, but it also assumed the genotoxic impurity would be as carcinogenic as some of the most potent carcinogens (Munro *et al.* 1999).

The purpose of this current work is to describe an extension of a risk-based approach for managing genotoxic metabolites of pharmaceuticals that relies on data generated during the normal progression of clinical development. The strategy is centered upon the availability of human ADME data at which point human metabolites are confirmed and quantitated. Using this information, human metabolites of genetic safety concern can be subsequently evaluated using a risk assessment model that identifies the probability that a known or suspect genotoxic metabolite will pose an excess cancer risk greater than 10^{-5} (i.e. 1 excess cancer over background per 100,000 people exposed). In addition, the results of the risk assessment will: 1) provide an understanding of the adequacy of metabolite representation in rodent carcinogenicity studies; 2) highlight the need to consider an alternative approach for carcinogenicity

testing; and, 3) inform potentially difficult drug development decisions. The rationale and approach used for each step of the risk assessment process is described herein. Representative case studies are used to illustrate the practicality of the approach within the context of drug development.

Genotoxic Hazard Identification of Metabolites Identified During Drug Development

Genetic toxicology testing is conducted in the early stages of drug development with the intent to identify hazards associated with both the parent molecule and its metabolites. This is accomplished by: (1) employing metabolic activation systems (typically Aroclor-induced rat liver S9) when conducting *in vitro* tests, and (2) conducting an *in vivo* genotoxicity assay—that is, the micronucleus assay—in rat or mouse. Many pharmaceutical sponsors conduct *in vitro* genetic toxicology screening assays during the drug discovery phase. In many cases, the observation of a genotoxic hazard in this early stage, whether under direct or metabolic conditions, will direct design efforts away from potentially hazardous substituents in favor of drug candidates without this potential liability. According to ICH M3 (ICH 1997d), all pharmaceutical sponsors must conduct standard *in vitro* genetic toxicology hazard identification studies such as Ames and mammalian cytogenetic assays prior to initiation of clinical investigations, whereby the *in vivo* micronucleus study is required prior to initiation of Phase 2 clinical investigations. Identification of a genotoxic hazard at this point in the drug development process may lead to further investigations to understand the relevance of the genotoxic response, or termination of development of the drug candidate. Attempting to understand the relevance of a positive result in the presence of Aroclor-induced S9 is not a simple task because the known *in vivo* metabolic profiles in rat and human may not be directly related to *in vitro* metabolism in the presence of S9.

The ability of *in vitro* and *in vivo* genotoxicity tests to model human metabolites of interest can be varied and therefore limit the capability to adequately conduct human

carcinogenicity risk assessment. The generation of both false positive and false negative results in genotoxicity tests is possible and the issues and potential shortcomings of induced rat liver S9 were recently reviewed in a report from the 4th International Workshop on Gentoxicity Testing (Ku *et al.* 2007). In addition, case examples were shared to illustrate other experimental approaches to address the shortcomings, such as the use of alternative metabolic activation systems and direct testing of metabolites of concern.

Given the limitations of standard genotoxicity testing methods, it seems prudent to consider the adequacy of human safety assessment of drug metabolites. Here we propose a risk assessment strategy triggered by human exposure to known or suspect genotoxic metabolites identified through the conduct of a human ADME study using radiolabeled drug, typically in Phase 2 of clinical development.

Risk Characterization Framework for Human Genotoxic Metabolites

The proposed framework for risk characterization of genotoxic human metabolites depends on three factors. The first is the probability that a compound with a structure-based concern for genotoxicity or data-driven evidence of genotoxicity could be a human carcinogen. *In silico* models and structural analysis are used to provide alerts for potentially positive results from genotoxicity assays (Matthews and Contrera 2007; Matthews *et al.* 2006b). Genotoxicity tests are used to evaluate positive findings from the models and screen compounds for their potential to be carcinogens. It is estimated that < 5-10% of all chemicals would be considered rodent carcinogens (Fung *et al.* 1995). However, compounds that result in positive genotoxicity tests are more closely associated with positive results from cancer studies in rodents. For example, the Ames assay has approximately a 60% concordance with results from rodent carcinogenicity studies (Matthews *et al.* 2006a). Yet the neoplasias developed in rodents can be species specific, with limited relevance to humans. So for this evaluation and others

(Barlow *et al.* 2001; Müller *et al.* 2006), it has been assumed that there is about a 50% probability that a genotoxic compound, determined by structure or appropriate hazard testing, could be carcinogenic to humans.

The second factor is the probability that a negligible cancer risk (i.e., 1 incidence in 100,000) will not be exceeded from exposure of the metabolite in humans. This can be determined using a risk distribution for chemicals known to be carcinogens. Each carcinogen in the risk distribution has an associated TD_{50}, or the average daily dose at which 50% of the population remains tumor free over a lifetime (Gaylor and Gold 1995; Gold *et al.* 1984; Gold *et al.* 1991; Peto *et al.* 1984; Sawyer *et al.* 1984). The estimated exposure to a genotoxic metabolite in humans was linearly extrapolated to a TD_{50}. That is, TD_{50} (mg/kg/day) = [human metabolite exposure (mg/kg/day) × (0.5/0.00001)], assuming the exposure was associated with a 1 in 100,000 excess cancer risk. The TD_{50} was compared relative to a distribution of known animal carcinogens [appendix 1.2 in (Fiori and Meyerhoff 2002)] to calculate the proportion of compounds that are greater than this TD_{50}. This proportion was considered to represent the probability of not exceeding an excess cancer risk of 1 in 100,000 from exposure of the metabolite in humans. There is precedent for using this level of risk for genotoxic impurities, and for carcinogenic residual solvents in pharmaceuticals (CHMP 2006; ICH 1997c). This approach is similar to risk assessments for impurities in food or pharmaceuticals where the genotoxic hazard is known (Fiori and Meyerhoff 2002; Kroes *et al.* 2004; Müller *et al.* 2006).

The final factor included in the risk characterization framework is the probability that, if a genotoxic metabolite was carcinogenic, then it would be detected as a positive tumorigenic response in animal studies. Risk assessments for genotoxic impurities have typically been based on limiting exposures so that a negligible human cancer risk would not be exceeded. As such, the level of the genotoxic impurity is reduced to low levels,

resulting in a limited ability to detect the tumorigenic effects of an impurity in the carcinogenicity study. Exposure to genotoxic chemicals from metabolism of drugs can occur, however, at a much higher total exposure level, which will result in an increased probability of detecting tumors via carcinogenicity studies in rodents. In this assessment, it was assumed that a 5% response is the limit of sensitivity for a typical carcinogenicity study (Cheeseman *et al.* 1999). This tumor response was extrapolated to a TD_{50}, based on the known total exposure to a genotoxic metabolite, using methods derived by Gaylor and Gold (1995), which are the same methods used to extrapolate TD_{50} values for known carcinogens in the Cancer Potency Database (http://potency.berkeley.edu/) (Gaylor and Gold 1995). The TD_{50} was then compared relative to a distribution of known animal carcinogens [Appendix 1.2 in (Fiori and Meyerhoff 2002)] to determine the proportion of compounds that have cancer potencies less than the TD_{50}. This proportion was then considered the probability of detecting tumors via carcinogenicity studies in rodents from exposure to the metabolite. Since this analysis is based on exposure of the metabolite, this takes into account the pharmacokinetic differences between species and the probability can be based on the maximum exposure in a single species.

Applying the three factors described above, the total probability of a genotoxic metabolite not exceeding a 1 in 100,000 excess cancer risk can be estimated. The probability of not exceeding an excess cancer risk of 1 in 100,000 from exposure of the metabolite in humans (PA) was added to the probability of detecting tumors via carcinogenicity studies in rodents (PB). This resultant probability was adjusted with the probability that a genotoxic compound is also carcinogenic (PC), with methodology similar to Munro *et al.* (1999). Thus, the total probability for not exceeding a 1 in 100,000 excess cancer risk was $(PA \cup PB) \cap (PC)$.

The total probability for not exceeding a 1 in 100,000 excess cancer risk was estimated for various combinations of animal and human doses ranging from 100 mg/kg to 0.00001 mg/kg (Table 1). A critical criterion for these estimations was that the 2-year oncogenicity study for the drug substance, the source of the genotoxic metabolite, did not result in a positive tumorigenic response explained by a genotoxic mode of action. It was assumed for the calculated probabilities that human exposure is lifetime daily exposure (70 years). Furthermore, exposure in rodents is assumed to be associated with standard 2-year rodent bioassays.

There were several generalizations derived from the analysis summarized in Figure 1. First, the probability of not exceeding a 10^{-5} excess cancer risk over background increases as exposure to rodents increases. Specifically, if the effective animal exposure to the metabolite is >10 mg/kg/day then there is a high probability (88-100%) of not exceeding the cancer risk. This is intuitive because high rodent exposures can increase detection in the carcinogenicity study. Second, the probability of not exceeding at 10^{-5} excess cancer risk over background increases as human exposure decreases. Low human exposure decreases the potential effect of a genotoxic metabolite. A metabolite exposure of 0.00001 mg/kg or less has a high probability (93-100%) of not exceeding the cancer risk. Lastly, probability also increases as the ratio of animal exposure over human exposure increases. When the ratio is 1000x or greater, there is a high probability (92-100%) of not exceeding the cancer risk.

Total Exposure-Based Risk Assessment for Human Metabolites

After a drug is administered to an organism, the number of possible metabolites that could be formed can be numerous. In most cases, the total number of metabolites observed is less than 100 and frequently in the range of 20 to 50. Drugs that possess a high lipophilicity tend to undergo multiple serial and parallel metabolic reactions before being converted to metabolites polar enough to be readily excreted. Metabolic pathways

for drugs that are more hydrophilic tend to be simpler. Furthermore, each drug can possess multiple potential sites for metabolism and even with modern analytical methodologies (e.g., mass spectrometry and NMR) it is not often straightforward to identify the structure of every metabolite. Thus, it is not possible to generate authentic standards of every potential metabolite of a new drug and to develop methods to detect and quantitate each of these entities. To circumvent this insurmountable task, studies are conducted as early as Phase 1 and as late as Phase 3 using radiolabeled analogs of drugs (typically carbon-14; sometimes tritium), in which the radiolabeled material is administered orally to laboratory animals and then to humans. Over the course of these studies, excreta and plasma are collected to assess mass balance and pharmacokinetics of the radioactivity. These matrices are usually analyzed by HPLC with radiometric detection to generate quantitative profiles of metabolites. Due to the complex data package that is generated, this is usually termed a human "ADME" study: absorption, distribution, metabolism, and excretion. Such a study is the only means by which a complete, comprehensive, and quantitative profile of metabolites of a new compound can be obtained. With modern spectroscopic techniques, such as mass spectrometry and NMR spectrometry, chemical structures for the metabolites can be proposed. Thus, in a radiometric ADME study, identification and quantitation of metabolites can be done in the absence of authentic standards or specific analytical methods for these metabolites. The quantitative metabolite profile data obtained in the human ADME study represents a powerful piece of information with which to make assessments of the suitability of preclinical safety information, including genotoxicity. However, it should be noted that for drugs in which the parent compound and/or drug-related material has a long half-life, the data obtained from the single dose ADME study may not be wholly representative of the exposures to metabolites upon repeated dosing. Pharmacokinetic modeling can be applied to make such estimates of steady-state

exposures, provided that the clearance does not change over time. This is a reasonable assumption for almost all cases.

In most ADME studies in humans, quantitative profiles of metabolites are obtained in plasma, urine, and fecal homogenates. Measurement of the exposures of various tissues to specific metabolites is not feasible in humans, and while such measurements could be made in small laboratory animal species, the data collection is of questionable value since there are no means by which to make a direct comparison to human. Knowledge of the circulating metabolite profile can provide insight into some specific types of toxicities (Smith and Obach 2005, 2006), but in general this information will be of lesser value in understanding the contribution of metabolites to risk of genotoxicity. Circulating metabolite profiles can be misleading when attempting to address the question of total systemic metabolite exposure. This is because the concentration of a metabolite in plasma is driven not only by how readily the metabolite is formed, but also by its distribution properties. Compounds that are highly bound to serum albumin, such as carboxylic acid metabolites, can represent very large peaks in HPLC radio-chromatograms from plasma, but because these do not partition readily outside the plasma compartment, their abundance in the entire body can be small. Other metabolites that tend to distribute into tissues, such as lipophilic amines, will appear to be minor in circulation, but can actually represent a large body burden. Also, typical cancer potency values (e.g. TD_{50}) are measured relative to dose (mg/kg/day) delivered to the animal and not plasma levels. For these reasons the metabolites in excreta were used in this analysis versus circulating metabolites.

The quantitative metabolite profile in excreta can be leveraged in gaining an understanding of the total exposure, albeit indirect, to metabolites. In this case, the exposure is one of a “total body burden” to a metabolite, and not a discrete concentration or AUC value. The body must be considered as one homogeneous

compartment, and the potential for a metabolite to selectively concentrate in a specific tissue cannot be factored in (or it is considered to be similar across species when making inter-species comparisons). To assess total body burden, the complete metabolic tree must be constructed from the excretory metabolite data, and intermediate metabolite structures may need to be inferred if they are not directly observed. (For example, if an excretory metabolite is a hydroxyglucuronide of the parent drug, and the hydroxylated intermediate metabolite itself is not observed in excreta, then its presence in the body must be inferred.) This is especially important for metabolites for which there is greater scrutiny on the possibility of genotoxicity due to chemically reactive structures. Such reactive metabolites are only rarely observed in radiometric HPLC profiles (due to instability and/or very low abundance), and it is their downstream stable end products which are observable. Common examples of such metabolite types include mercapturic acids (arising via glutathione conjugation of reactive electrophiles), dihydrodiols (arising from hydrolysis of epoxides), and lactams (arising via oxidation of iminium ions) (Kalgutkar *et al.* 2005). An example of using metabolite profile data in this manner is shown in Figure 2 and in Table 2.

In this example, a hypothetical drug is metabolized in humans via three initial pathways and these initial metabolites are further metabolized and/or excreted. The total number of metabolites observed is seven (M1, M3, M5, M6, M7, M8, M9) from a total of 10 reaction pathways (some of which have more than one reaction within the arrow). The structures of two metabolites not observed (M2 and M4) must be inferred due to the presence of their downstream metabolites. For the terminal metabolites (M5, M6, M7, M8, and M9), the estimation of total body burden is straightforward and is merely the total dose times the percentage that each represents. However, for the intermediate metabolites, an estimate of the body burden comprises not only how much was directly observed, but also how much was converted to downstream products. Thus

for M3, this is not only the 5 mg equivalents observed in urine, but also the 20 mg equivalents comprised by M8 and M9, which had to arise via M3.

For potential parallel pathways, an estimate requires including both routes, even though the precise route cannot be delineated. This is illustrated by M1. The estimate of body burden for M1 must include not only the downstream metabolites clearly attributed to arise via M1 (i.e. M6 and M7), but also the metabolite that could have arisen partially or wholly from a parallel pathway (M5). Because it cannot be known how much M5 arose via the M1 versus M2 pathways, it can be assumed that all M5 came from M1 for the purpose of making an estimate of body burden for M1. And conversely, when estimating body burden for M2, it must also be assumed that all of M5 arose via M2. Finally, what is perhaps most relevant for assessing the potential for genotoxicity in this particular example is the estimation of the body burden of M4, a metabolite closely related to a reactive intermediate (quinone). The observation of a mercapturic acid in excreta indicates exposure to a reactive electrophilic quinone which was not directly observed. To make an estimate of the body burden to such a reactive intermediate, the metabolites deriving from the pathway must be summed, in this case M6 and M7 for a total of 35 mg equivalents. This likely represents an overestimate, since the formation of M7, a glucuronide conjugate of the hydroquinone, represents a detoxification pathway that could reduce the amount of quinone formed from M4. However, since hydroquinones and quinones are known to be readily inter-convertible through redox cycling, assuming that the body burden to quinone is equivalent to that of the hydroquinone would be most appropriate.

Upon completion of this exercise from the radiolabel ADME data for both humans and rodents, total body burden estimates for metabolites of genotoxic concern are available for both species. This information can then facilitate a risk assessment. That is, the body burden estimates for human and rodent are considered relative to the

generalizations defined in Figure 1. In addition, for each metabolite, the probability of exceeding an excess cancer risk of 10^{-5} can be calculated taking into consideration the 3 factors described previously (see Risk Characterization for Human Genotoxic Metabolites).

Case Studies

Following, are three case examples that exemplify the application of this framework for the risk assessment of human genotoxic metabolites. In the case studies it is assumed that the % of metabolite formed is independent of dose and duration of treatment. In these case studies a 70 kg person, linear pharmacokinetics and complete recovery of the administered dose was assumed.

Case Study #1

An active pharmaceutical ingredient intended for the treatment of bipolar disorder is in Phase 2 clinical development. An intermediate used in the synthesis of the active pharmaceutical ingredient, which is also a metabolite in human, is tested in the Ames assay and the *in vitro* cytogenetics assay using human lymphocytes (HLA) to comply with worker safety and transportation regulations. This shared intermediate/metabolite (M5) produces positive responses in both the Ames and HLA assay. Furthermore, two human metabolites downstream of M5 (that is, M2 and M1) are also suspect genotoxic metabolites based on high structural similarity to M5. The metabolic pathway is sequential: parent → M5 → M2 →M1.

Based on the projected human efficacious dose (160 mg) and the highest dose utilized in 2 year carcinogenicity testing in mouse and rat (200 mg/kg and 12 mg/kg), the maximum total body burden to M5, M2 and M1 were estimated for each species (human, rat and mouse). The body burden of M5 was estimated taking into consideration the % of M5 detected in excreta plus the % M1 and M2, which are direct downstream metabolites. For example, in rat the percentages of M5, M1 and M2 were 7.3, 5.1 and

1.1%, respectively, therefore the body burden of M5 in rat is estimated at 13.5%. Similarly the body burden of M2 was estimated by adding the % of M2 detected plus the % of M1 detected. The total percentage of metabolite burden, was then multiplied by the highest dose to which each species is exposed. The results of these tabulations are presented in Table 3. An example illustrating how the rat total body burden was calculated follows:

Total Rat Body Burden of M5 (mg/kg) =

(% M5 + %M2 + %M1) × (Highest Dose in Carcinogenicity Study)

= (1.1% + 5.1% +7.3%) ×(12 mg/kg/day)

=(13.5%) × (12 mg/kg/day)

= 1.620 mg/kg/day

After body burden estimates were determined for each metabolite of concern for each species, an assessment of risk was conducted by considering the generalizations established from the risk characterization model (Figure 1). In addition, for each metabolite the probability of not exceeding an excess cancer risk of 10^{-5} was calculated. For metabolites, M5, M2 and M1, the total burden of exposure in human is 0.286, 0.251 and 0.251 mg/kg/day, respectively (Table 3). In the case of M5, the total body burden in rat and mouse at the highest dose used in 2-year carcinogenicity testing is estimated at 1.620 and 49.6 mg/kg respectively. For M2, the exposure estimates are 0.744 and 46.4 mg/kg, respectively for rat and mouse. For M1, the exposure estimates are 0.132 and 38.0 mg/kg, respectively. At this point it should be noted that only the mouse has exposures for the three metabolites that are >10 mg/kg/day. Therefore, the maximum ratio of animal (i.e., mouse) exposure over human exposure is 173X, 185X and 151X for M5, M2 and M1 respectively. Using these estimates of total body burden to M5, M2 and M1 the probability of controlling excess cancer risk to 10^{-5} was estimated to be ~95% based on a dose-specific calculation in a relevant rodent species. In this case, human

exposures were well in excess of 1.5 µg/day, while exposures in the rat carcinogenicity study were less than 10 mg/kg/day. The high probability of not exceeding an excess cancer risk of 10^{-5} was largely related to the high exposures achieved in the mouse carcinogenicity studies, which provides a high degree of confidence in the ability to qualify the carcinogenic potential of these metabolites in the non-clinical testing.

Case Study #2

An active pharmaceutical ingredient is in Phase 2 clinical development for the treatment of hypercholesterolemia. The human ADME study revealed two simple quinoline metabolites, M3 and M4 (Figure 3). Both M3 and M4 were also observed in rat and mouse ADME studies.

Based on the projected human efficacious dose (60 mg) and the highest dose utilized in 2 year carcinogenicity testing in mouse (300 mg/kg/day) and rat (400 mg/kg/day), estimates of the maximum total body burden to M3 and M4 were determined for each species (human, rat and mouse). The body burden of M3 was estimated taking into consideration the % of M3 detected in excreta plus the % metabolite for all metabolites in excreta which are directly downstream of M3. This total percentage of metabolite burden, which represents M3+M4+M5+M21+M22, was then multiplied by the highest dose to which each species would be exposed. After body burden estimates were determined for each metabolite of concern for each species, an assessment of risk was conducted by considering the generalizations established from the risk characterization model (Figure 1). In addition, for each metabolite the probability of not exceeding an excess cancer risk of 10^{-5} was calculated.

For M3 and M4, the total burden of exposure in human was estimated to be 0.164 and 0.102 mg/kg/day, respectively (Table 4). In the case of M3, the total body burden in rat and mouse at the highest dose used in 2-year oncogenicity testing is estimated at 90.4 and 88.5 mg/kg respectively. For M4, the exposure estimates are 90.4

and 77.8 mg/kg, respectively for rat and mouse. Therefore, the maximum ratio of animal exposure over human exposure is 550X and 883X for M3 and M4 respectively. Using the estimates of total body burden to M3 and M4 the probability of controlling excess cancer risk to 10^{-5} was determined to be >95%. In this case, given that human exposures were well in excess of 1.5 μg/day, the high probability of not exceeding an excess cancer risk of 10^{-5} was primarily related to the high exposures achieved in the rat and mouse carcinogenicity studies, which provide a high degree of confidence in the ability to qualify the carcinogenic potential of these metabolites in the non-clinical testing.

Case Study #3

An active pharmaceutical ingredient intended for the treatment of GERD is in Phase 2 clinical development. A human metabolite (M22) which had previously been assigned as an amine metabolite was determined to be a nitroalkane metabolite (Figure 4). This structurally alerting metabolite represented 10% of excreted drug in human and was observed in substantially lower quantities in mouse excreta (2.3%) and was not detected in rat. Based on the projected human efficacious dose (10 mg) and the highest dose intended for the 2 year carcinogenicity testing in mouse (100 mg/kg), estimates of the maximum total body burden of M22 were estimated for each species (human and mouse). The body burden of M22 was estimated by taking into consideration the % of M22 detected in excreta. There were no downstream metabolites of M22 identified in either human or mouse. The total percentage of metabolite burden, was then multiplied by the highest dose to which each species will be exposed. The results of these tabulations are presented in Table 5.

For M22 the total body burden of exposure in human and mouse is 0.014 mg/kg and 2.3 mg/kg respectively (Table 5). Therefore, the maximum ratio of animal exposure over human exposure is estimated to be ~164X. Based on the body burden estimates an assessment of risk was conducted by considering the generalizations established

from the risk characterization model (Figure 1). In addition, the probability of not exceeding an excess cancer risk of 10^{-5} was calculated. In this case study, none of the criteria established in the generalizations were met. That is, human exposures were in excess of 1.5 µg/day, exposure in the mouse oncogenicity study was less than 10 mg/kg/day, and there was less than 1000X margin of animal exposure to human exposure. Accordingly, in this case the probability of not exceeding an excess cancer risk of 10^{-5} was only 84% as opposed to the prior two cases in which at least one risk criterion was met.

In this case study, there was a lower probability of not exceeding a 10^{-5} excess cancer risk. The model developed in this paper may not be sufficient to demonstrate safety of the metabolite. However, this model should be viewed as one of many tools for assessing metabolite safety. More information could be gathered about the metabolite to make a "weight of evidence" determination of its genotoxic potential. Additional safety studies on a case-by-case basis could be performed directly with the metabolite or the metabolite could be "spiked" into the dose solution to ensure adequate exposure. However, it should be noted that there are often separate ADME issues associated with direct administration of a metabolite, since the disposition of the metabolite may be vastly different when it is administered directly versus when it is generated from a parent drug *in situ*. Thus, other risk-benefit assessment approaches must be considered as well.

Discussion

The role of metabolism in the generation of reactive electrophiles that are responsible for forming covalent bonds to nucleic acids and causing mutation and ultimately carcinogenicity is well established for many chemicals. Therefore, most pharmaceutical sponsors start working in the early stages of drug discovery to predict and avoid the potential for the formation of highly reactive and possibly genotoxic

metabolites in vivo. In addition, most sponsors employ genotoxicity screening assays using a metabolic activation system (i.e. Aroclor induced rat liver S9) to identify drug candidates with potential genotoxic metabolites. The ICH guidance requires the conduct of an in vivo genotoxicity test and *in vitro* genotoxicity tests with an activation system (ICH 1996, 1997c) with the intent to assess the genotoxic potential of the active pharmaceutical ingredient and its respective metabolites prior to initiating clinical development.

Despite these efforts, the ability of both *in vitro* and in vivo test systems to model human metabolites of interest can be variable and thus limit the ability to adequately conduct genotoxicity and carcinogenicity risk assessment in certain cases. Recently an International Workshop of Genotoxicity Tests (IWGT) focused on the limitations of the current approach to genotoxicity testing, some of which likely account for the failure of the standard test systems to detect several suspect human carcinogens (Ku *et al.* 2007). Important limitations of *in vitro* test systems that were highlighted in the output from the workshop include the fact that non-CYP enzymes may not be active due to absence of necessary cofactors and that co-factor supplementation may not overcome the issue due to lack of cell permeability (Glatt 2000; Glatt and Meinl 2005; Mulder *et al.* 1977). That is, reactive metabolite formation may only occur extracellularly and therefore limit the ability to react with DNA. It was also noted that CYP enzymes from rat and human differ in their substrate specificity and the reactions they catalyze. Furthermore, it is important to note that Aroclor treatment of rats used to prepare S9 creates an unnatural complement of CYP enzymes, with high induction of CYP1A and B, and minor increasing or decreasing effects on other CYPs (Guengerich *et al.* 1982). These factors are likely to contribute to differences observed in the qualitative and quantitative profiles of drug metabolites formed in studies comparing human liver S9 to Aroclor-induced rat S9 (Obach and Dobo 2008). Therefore, multiple factors may limit *in vitro* hazard

identification of genotoxic metabolites. In vivo, the limitations of the assay are likely related to the pharmacokinetics of the test compound and its possible species-, sex- and tissue-specificity.

Despite the rigorous assessment of genetic safety required to support clinical development, given the limitations cited above, it seems prudent to consider the adequacy of the testing as information regarding human metabolites becomes available. A number of publications have focused on recommendations for a framework for assessing the safety of metabolites (Baillie *et al.* 2002; Smith and Obach 2005). Most recently the FDA issued a guidance document (USFDA 2008b) and several pharmaceutical sponsors have outlined their approaches to addressing this important issue (Humphreys and Unger 2006; Luffer-Atlas 2008). It should be noted that the FDA guidance does provide recommendations concerning genotoxicity testing of human metabolites that exceed a 10% relative abundance threshold (USFDA 2008b). To date there has been no framework recommended to facilitate a risk assessment for confirmed human genotoxic metabolites. Since the identification of such metabolites is particularly challenging when realized in the later stages of drug development, it is desirable to have a framework for conducting a carcinogenic risk assessment. Here we propose a risk assessment strategy for human genotoxic metabolites that relies on data generated during the conduct of rodent and human radiolabeled ADME studies. The radiolabeled ADME studies provide a complete, comprehensive, and quantitative profile of metabolites, which allows the total body burden of any given metabolite to be estimated. This absolute exposure data is necessary to support the cancer risk assessment for genotoxic metabolites. That is, any metabolite of genotoxic concern confirmed to be present in both humans and rodents can be evaluated using this model to determine if there is a favorable assessment of human risk. While this document may provide a model for carcinogenic outcomes of genotoxic metabolites, it does not address other

toxicities that may be associated with a genotoxic species such as reproductive toxicity. Therefore, this model should only be used to address potential carcinogenicity and other toxicities should also be addressed by the available guidances.

The method proposed incorporates aspects of prior approaches for impurities in food or pharmaceuticals where the genotoxic hazard is solely known for the impurity, i.e. there is no data on carcinogenicity (Fiori and Meyerhoff 2002; Kroes *et al.* 2004; Müller *et al.* 2006). The risk assessment/management approaches for genotoxic compounds are difficult to validate given the low acceptable occurrence for the effect (e.g. 1 in 100,000 excess cancer risk) and the limited data available (e.g. Ames assay results). Therefore a conservative philosophy is typically applied given the uncertainties in the outcome. In general, the probability of exceeding an excess cancer risk of 10^{-5} is estimated by referencing the cancer potency curve for rodent carcinogenicity studies, which was derived by simple linear extrapolation from carcinogenic potencies giving a 50% tumor incidence (TD_{50}) over background for more than 700 carcinogens (Fiori and Meyerhoff 2002; Kroes *et al.* 2004; Müller *et al.* 2006). Several generalizations were defined to facilitate the risk assessment of case studies, including an acceptable absolute exposure limit of 1.5 μg/day for lifetime exposure in humans (see Table 1 and 2). The 1.5 μg daily dose is consistent with the limit established for lifetime daily intake of a genotoxic impurity in active pharmaceutical ingredients (CHMP 2006; Müller *et al.* 2006). It is important to note that the derivation of the limit is dependent on conservative assumptions including no threshold for toxicity, the shape of the dose-response curve for all genotoxic carcinogens is linear, and the genotoxic compound could be a highly potent carcinogen. Data used to extrapolate to a risk associated with 10^{-5} is based on the most sensitive species, and on the most sensitive target tissue.

Furthermore, the daily exposure may need to be adjusted for those pharmaceuticals administered over a less than lifetime duration. In the case of

genotoxic impurities it has been recommended that the duration of treatment be accounted for in establishing an allowable limit, that is staged TTCs have been defined for short-term exposure conditions (Müller *et al.* 2006). This is because carcinogenicity is a result of an accumulation of exposure over a lifetime. Although all of the case studies presented are drugs intended for chronic use, a staged TTC approach could be incorporated into the risk assessment framework for genotoxic metabolites to account for treatment scenarios that are less than chronic (70 years) duration. This would increase the probability of not exceeding an excess cancer risk of 1 in 100,000 in humans (PA) for a given exposure. Thus an adjustment could be made on Table 1 to reflect the change from long-term to short-term exposure to the genotoxic metabolite.

In the evaluation of these case studies, it became apparent that an exposure of 1.5 µg/day, associated with an excess cancer risk of 10^{-5}, is not a practical limit for metabolites. In none of the industry case studies evaluated to date, including those presented herein, have metabolites present at such a low total body burden been identified and quantitated. This is associated with the fact that radiometric HPLC, which is used to quantitate metabolites, struggles to reliably quantitate metabolites that represent less than 5% of the radiolabeled material (Smith and Obach 2005). Therefore, even for low dose drugs, 10 or 1 mg/day, the limit for quantitating metabolites would be 500 or 50 µg/day, respectively. Due to this limitation, the model as proposed would not be useful for risk assessment of unique human metabolites.

The exposure limit of 1.5 µg/day was selected as an extension of the approach recommended for genotoxic impurities; however, there are differences in relation to genotoxic metabolites which may allow for a different risk management outlook. In the case of genotoxic impurities in active pharmaceutical ingredients the acceptable excess cancer risk is exceptionally low because the impurities in the drug product provide no benefit to the patient. In contrast, drug metabolism is associated with desirable drug

properties such as pharmacokinetics, pharmacology and clearance. Also, exposure to the metabolite cannot be reduced for a given drug substance whereas strategies in manufacturing can be developed that can effectively reduce impurity levels. Therefore exposure to the metabolite is part of the risk/benefit assessment for each molecule.

Unlike other risk assessments, the proposed method for genotoxic metabolites takes advantage of the possibility that exposures in rodent carcinogenicity testing may be relatively high compared to human exposure. This additional factor, which is the assessment of total body burden in rodents, was incorporated into the risk assessment framework to assess the probability that the metabolite exposure in a 2-year rat or mouse carcinogenicity assay is high enough to detect an oncogenic and/or tumorigenic signal. Based on our initial evaluation of case studies, it appears that the total body burden to genotoxic metabolites in rodents has a greater influence on the outcome of the risk assessment than other factors. Therefore, it seems that the useful aspect of this methodology is that it allows one to understand if the amount of metabolite formed in the rodent is sufficient to have confidence that a negative outcome in a carcinogenicity study in essence “qualifies” the metabolite(s) of concern.

This model can be most useful for study design of a carcinogenicity study if the genotoxic metabolite was discovered early in development. This timing makes it possible that the model can be used proactively when selecting doses for the cancer bioassays. That is, assuming that there is knowledge of a human metabolite of genotoxic concern prior to initiation of the study and that the metabolite is also formed in rodent, then a sponsor could select doses that would provide confidence in metabolite qualification (e.g. ≥ 10 mg/kg/day exposure to metabolite or other criteria). Alternatively, if the model is not well defined prior to initiation of the carcinogenicity studies, it is possible that a key metabolite might not be adequately addressed and the development program will carry a risk of needing to do additional studies after submission.

The described model has limitations which may diminish its application in some cases. For example, if a metabolite of concern is discovered at low levels in animals and humans (1-5%), the model may indicate a concern when none exists. This is because in practice the high dose in animals is more likely to result in a favorable outcome than a low dose in humans. In most instances one will not be able to detect levels as low as 1.5 µg/day in excreted material. However, low human exposure levels of metabolites are typically not of concern and therefore the model should be applied when an issue is triggered by application of the available regulatory guidances. Also, when a metabolite is related to a drug for an unmet medical need, one must also consider the benefit of the medication when considering the risks of the metabolite. A precedent for this perspective can be found in the FDA Guidance on Safety Testing of Drug Metabolites, which exempts drugs for life-saving diseases from the proposed recommendations (Smith and Obach 2005). Furthermore, the patient population is another consideration and increased conservatism may be applied to certain groups that are more susceptible to cancer.

There are also several assumptions that a person should be aware of when applying the model. Many of the assumptions are inherent to the model as described in the methods such as the probability a genotoxic compound is also carcinogenic or the sensitivity of a carcinogenicity study to detect a tumor. Another assumption is the metabolite once formed will penetrate the target tissue before it is further metabolized. Furthermore, the carcinogenicity potency database from which the model was based on contains carcinogenic compounds that are nongenotoxic. If genotoxic compounds are more potent carcinogens than nongenotoxic compounds resulting in a higher TD_{50} for nongenotoxic carcinogens, then this may overestimate the upper dose needed to qualify a genotoxic metabolite. Thus a lower dose may be appropriate to qualify genotoxic metabolites in animals. Furthermore, the model was developed from known carcinogens

which undergo their own metabolism. The mechanism of action for these carcinogens is in many instances based on their own genotoxic metabolites. A more accurate description would use the body burden of the genotoxic metabolite for each carcinogen, but this dataset is limited. This model should be taken in context of risk management, where despite its limitations, is still a good model for making decisions in certain instances. Therefore, good judgment, considerations of other guidances, and understanding the risk/benefit of the medication is essential before use of the model.

While this model derives an acceptable margin of safety (e.g. 1000x), it is noteworthy to highlight the differences between its assumptions and guidance typically used in pharmaceutical development to derive a safe starting dose in volunteers (USFDA 2005). The USFDA guidance uses NOAELs derived from nonclinical species for the drug substance and converts these levels to human equivalency doses (HEDs) by taking into account allometric scaling differences between animals and humans. A safety factor is then applied to the HED to derive a safe human dose and allow for an appropriate margin of safety. In this case, there is an assumption of a threshold, which allows for a smaller margin. A probabilistic approach was applied to genotoxic metabolites in a similar fashion as the TTC. Allometric scaling was not applied to the derivation of the TTC because it was risk-based and a large degree of conservative assumptions were already made in its derivation (Kroes *et al.* 2004; Munro *et al.* 1999). While some of the generalizations for genotoxic metabolites expand upon the TTC used for impurities, there was no reason to differ in this assumption as well.

In summary, a framework has been proposed which may be useful for conducting a risk assessment of human genotoxic metabolites. The method, as it has been applied here, is intended to be used in the later stages of development when exposures to metabolites are quantitated through radiolabel ADME studies in both human and rodents. The approach is practical in that it relies on data that is required to

be generated during the normal process of drug development. The assessment of case studies to date suggests that the conservative nature of the model makes it impractical for conducting a risk assessment for any unique human genotoxic metabolites. The TTC was developed for impurities, which is based solely on developing a dose that would result in a negligible risk to humans for most compounds if they turned out to be carcinogenic. Solely applying the TTC for metabolites of genotoxic concern is not only impractical but it also does not take into account that metabolites are part of the risk/benefit equation. In cases wherein metabolites of concern are present in rodents, the risk assessment is useful, providing an understanding of the adequacy of metabolite representation in rodent carcinogenicity studies. This model may be most useful for managing the challenges associated with identification of a genotoxic human metabolite in the later stages of drug development. It may also be useful for non-clinical study design from early discovery of a genotoxic metabolite.

Tables

Table 1. Total probability of not exceeding a 1 in 100,000 cancer risk for a genotoxic metabolite.

		Human Exposure to Metabolite (mg/kg)							
Animal Preclinical Exposure to Metabolite (mg/kg)		100	10	1	0.1	0.01	0.001	0.0001	0.00001
	100	97%	97%	97%	97%	100%	100%	100%	100%
	10	88%	88%	88%	88%	94%	100%	100%	100%
	1	74%	74%	74%	74%	80%	92%	100%	100%
	0.1	61%	61%	61%	61%	67%	78%	93%	100%
	0.01	54%	54%	54%	54%	60%	71%	86%	97%
	0.001	51%	51%	51%	51%	57%	69%	83%	94%
	0.0001	50%	50%	50%	50%	56%	68%	83%	93%
	0.00001	50%	50%	50%	50%	56%	68%	82%	93%

a. The model was adapted from Fiori and Meyerhoff (2002) which analyzed the carcinogenic potencies of 705 animal carcinogens.

b. Animal preclinical exposure is the total body burden to the metabolite estimated in the 2 year oncogenicity study. It can be estimated by using short-term radiolabeled animal studies in the same species.

c. Human exposure is the total body burden to the metabolite estimated in humans. It can be estimated by using radiolabeled clinical studies.

d. ▢ Where combination of exposure in animals and humans results in a probability below 85 percent of not exceeding a 1 in 100,000 excess risk of cancer.

Table 2. Calculated body burden estimates for metabolites for a hypothetical example (Refer to the accompanying metabolic tree in Figure 2).

Metabolite	**% in Excreta**	**Sum of:**	**% of Dose**	**Body Burden**
Phenol M1	20	M1+M4+M6+M7+M5	70	70 mg eq.
N-Desmethyl M2	0	M5	15	15 mg eq.
Hydroxy M3	5	M3+M8+M9	25	25 mg eq.
Hydroquinone M4	0	M6+M7	35	35 mg eq.
N-Desmethyl Phenol M5	15	M5	15	15 mg eq.
Mercpaturic Acid M6	5	M6	5	5 mg eq.
Glucuronide M7	30	M7	30	30 mg eq.
Sulfate M8	5	M8	5	5 mg eq.
Dihydroxy M9	15	M9	15	15 mg eq.

Table 3. Case Study 1 - Estimates of Total Body Burden to Genotoxic Metabolites in Human and Rodent.

Metabolite	**Human %**[a]	**Human Body Burden (mg/kg)**[b]	**Rat %**[a]	**Rat Body Burden (mg/kg)**[c]	**Mouse %**[a]	**Mouse Body Burden (mg/kg)**[c]
Parent	3.3	-	5.9	-	44.0	-
M1	11.0[d]	0.251	1.1	0.132	19.0	38.0
M2	11.0[d]	0.251	6.2	0.744	23.2	46.4
M5	12.5	0.286	13.5	1.620	24.8	49.6

a. The % of each metabolite was estimated by adding the % of the metabolite detected in excreta plus the % of its respective downstream metabolites detected in excreta.

b. The human body burden for each metabolite was estimated by multiplying the % of metabolite by the human efficacious dose (160 mg and assuming a 70 kg person).

c. The rodent body burden for each metabolite was estimated by multiplying the % of metabolite by the maximum daily dose used in the carcinogenicity study (200 mg/kg in mouse and 12 mg/kg in rat).

d. In the human ADME study M1 and M2 co-eluted, and represented a total of 11% of metabolite in excreta.

Table 4. Case Study 2 - Estimates of Total Body Burden to Genotoxic Metabolites in Human and Rodent.

Metabolite	**Human %[a]**	**Human Body Burden (mg/kg)[b]**	**Rat %[a]**	**Rat Body Burden (mg/kg)[c]**	**Mouse %[a]**	**Mouse Body Burden (mg/kg)[c]**
Parent	5.8	-	39.0	-	36.0	-
M3	19.1	0.164	22.6	90.4	29.5	88.5
M4	11.9	0.102	22.6	90.4	26.0	78.0

a. The % of each metabolite was estimated by adding the % of the metabolite detected in excreta plus the % of its respective downstream metabolites detected in excreta.

b. The human body burden for each metabolite was estimated by multiplying the % of metabolite by the human efficacious dose (60 mg and assuming a 70 kg person).

c. The rodent body burden for each metabolite was estimated by multiplying the % of metabolite by the maximum daily dose used in the carcinogenicity study (300 mg/kg/day in mouse and 400 mg/kg/day in rat).

Table 5. Case Study 3 - Estimates of Total Body Burden to Genotoxic Metabolites in Human and Rodent.

Metabolite	**Human %**[a]	**Human Body Burden (mg/kg)**[b]	**Rat %**[a]	**Rat Body Burden (mg/kg)**[c]	**Mouse %**[a]	**Mouse Body Burden (mg/kg)**[c]
Parent	9.0	-	29.0	-	34	-
M22	10.0	0.014	ND[d]	-	2.3	2.3

a. The % of each metabolite was estimated by adding the % of the metabolite detected in excreta plus the % of its respective downstream metabolites detected in excreta.

b. The human body burden for each metabolite was estimated by multiplying the % of metabolite by the human efficacious dose (10 mg and assuming a 70 kg person).

c. The rodent body burden for each metabolite was estimated by multiplying the % of metabolite by the maximum daily dose used in the carcinogenicity study (100 mg/kg/day in mouse).

d. None detected.

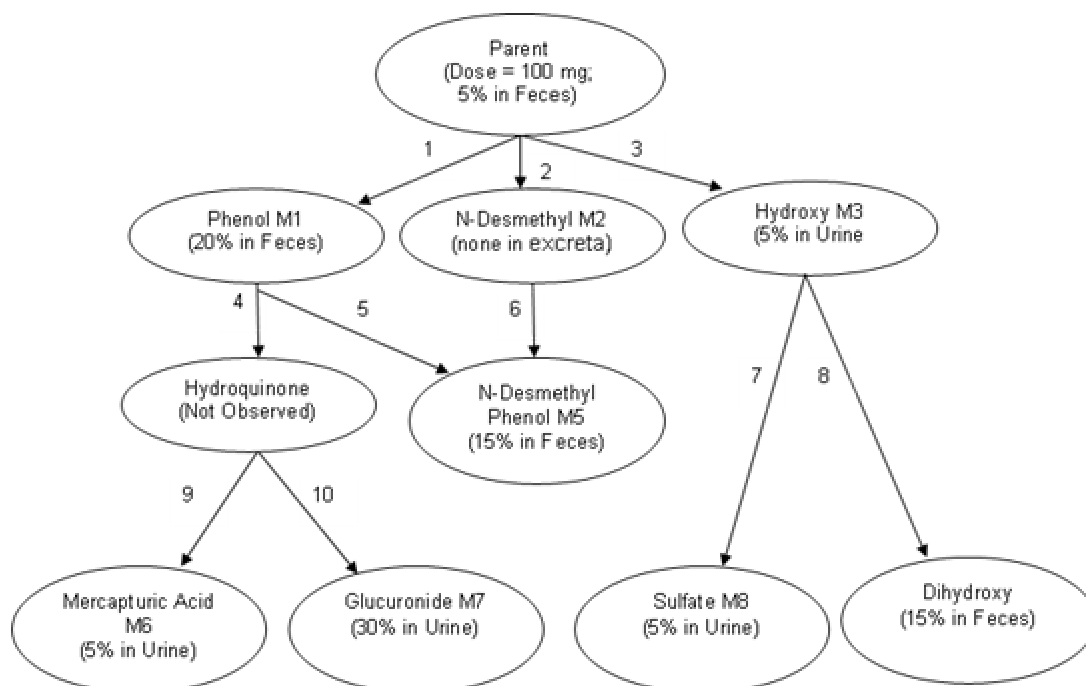
Figures

Metabolite is Considered Potentially Genotoxic

- Using radiolabeled studies, what is the internal dose in humans and in rodents (rats, mice or species used in carcinogenicity study)? This can estimate total body burden.
- Is the total body burden in rodents ≥ 10 mg/kg? Yes – there is a high likelihood of detecting a tumorigenic response in the carcinogenicity study.
- Is the total body burden in humans ≤ 1.5 µg/day? Yes – there is a high likelihood that human risk is below 1 in 100,000.
- Does the rodent total body burden exceed human's by ≥ 1,000x? Yes – that combination has a high probability of not exceeding a 1 in 100,000 excess cancer risk.
- What is the calculated probability of not exceeding a 1 in 100,000? excess cancer risk? This can help with risk management decisions about the compound.

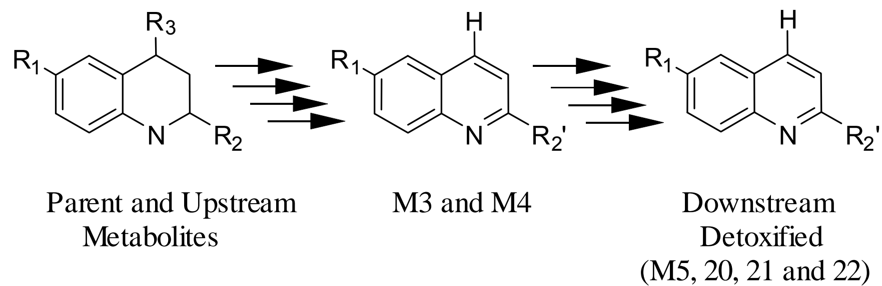
Note: These questions are intended solely as potential guides. If the data does not fit the particular questions, then the developed model is unlikely to be applicable. Other criteria may be used to show acceptability of a genotoxic metabolite.

Figure 1. Generalizations / Applications of the Risk Assessment Model for Human Genotoxic Metabolites.



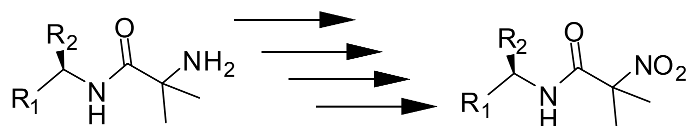
The parent compound is cleared by three initial pathways (1, 2, and 3) leading to three metabolites, one which is inferred as a metabolic intermediate (M2) because it is not observed in excreta. Subsequent biotransformation reactions lead to the metabolites that ultimately are excreted. Refer to accompanying Table 2 for calculations of metabolite body burden.

Figure 2. Hypothetical example of a metabolic tree constructed from in vivo excretory metabolite data.



The parent and metabolites upstream of M3 and M4 are non-alerting.
M3 and M4 are simple structurally alerting quinoline metabolites.
All metabolites downstream of M3 and M4 quinoline (i.e. M5, M20, M21 and M22) are non-alerting detoxified metabolites.

Figure 3. Case Study 2 - Representation of human metabolic tree.



Parent and Upstream Metabolites

Final Downstream Metabolite (M22)

In the case of case study 3, the final downstream metabolite observed in human excreta (M22) is structurally alerting, containing a tertiary nitro group.

Figure 4. Case Study 3 - Representation of human metabolic tree.

TOXICOGENOMICS AND CANCER RISK ASSESSMENT: A FRAMEWORK FOR KEY EVENT ANALYSIS AND DOSE-RESPONSE ASSESSMENT FOR NONGENOTOXIC CARCINOGENS[1]

Abstract

In order to determine a threshold for nongenotoxic carcinogens, the traditional risk assessment approach has been to identify a mode of action (MOA) with a nonlinear dose response. The dose-response for one or more key event(s) linked to the MOA for carcinogenicity allows a point of departure (POD) to be selected from the most sensitive effect-dose or no-effect dose. However, this can be challenging because multiple MOAs and key events may exist for carcinogenicity and oftentimes extensive research is required to elucidate the MOA. In this study, a microarray analysis was conducted to determine if a POD could be identified following short-term oral rat exposure with two nongenotoxic rodent carcinogens, fenofibrate and methapyrilene, using a benchmark dose analysis of genes aggregated in Kegg pathways and Gene Ontology (GO) biological processes, which likely encompass key event(s) for carcinogenicity. The gene expression response for fenofibrate given to rats for 2 days was consistent with its MOA and known key events linked to PPARα activation. The temporal response from daily dosing with methapyrilene demonstrated biological complexity with waves of pathways/biological processes occurring over 1, 3, and 7 days; nonetheless, the benchmark dose values were consistent over time. When comparing the dose-response of toxicogenomic data to tumorigenesis or precursor events, the toxicogenomics POD was slightly below any effect-level. Our results suggest that toxicogenomic analysis using short-term studies can be used to identify a threshold for nongenotoxic

[1] This chapter has been submitted to be published in: Bercu JP, Jolly RA, Flagella KM, Baker TK, Romero P, and Stevens JL. (2010). Toxicogenomics and cancer risk assessment: A framework for key event analysis and dose-response assessment for nongenotoxic carcinogens. Toxicol. Sci. (Submitted).

carcinogens based on evaluation of potential key event(s) which then can be used within a risk assessment framework.

Introduction

Cancer risk assessment is a process through which the risk of developing cancer over background can be determined following environmental exposure to a chemical (USEPA 2005). Cancer dose-response assessment depends on one major assumption, whether or not there is a threshold-effect. If no threshold is identified, it is assumed that a carcinogenic response can occur at any exposure (USEPA 2005). This type of assessment utilizes a linear low dose risk assessment model from which excess risk of cancer over background is extrapolated to low dose levels from effects occurring in animals at high doses (Gold *et al.* 2003) or a large margin of exposure from a tumorigenic dose (Barlow *et al.* 2006). If a threshold is assumed, the cancer risk assessment process typically involves identifying a point of departure (POD) from which an exposure level that would not cause an appreciable adverse effect (tumorigenesis) is estimated. The POD is typically derived from a no-observed effect level (NOEL), or a lowest-observed effect level (LOEL) if a NOEL cannot be identified, and adjustments (i.e. uncertainty factors) are made to lower the POD to an acceptable dose (Barnes and Dourson 1988; Dourson 1993; USEPA 2002).

Nongenotoxic carcinogens are a class of carcinogens where the assumption of a threshold seems applicable (Klaunig *et al.* 2000). According to the 2005 USEPA carcinogenicity risk assessment guidelines, a threshold response can be assumed for a nongenotoxic carcinogen if a threshold-based mode of action (MOA) can be described (USEPA 2005). The MOA approach does not imply a complete understanding of mechanism or a detailed description of all events linked to the mechanism of carcinogenesis at the molecular level. However, establishing a MOA does require description of one or more critical key event(s) necessary to cause cancer (Boobis *et al.*

2006). The assumption is that the no-effect dose for a key event, often defined by a biochemical or morphological endpoint, can be used to establish a threshold below which development of cancer is unlikely. While useful, this approach has not been applied extensively for rodent carcinogenicity data. Elucidating the MOA and identifying key events experimentally is time-consuming since only one or a few potential key events can be tested at a time and a compound may have multiple MOAs each of which may have multiple key events.

Regardless of the approach to risk assessment, it is important to identify a dose-response relationship to establish a threshold. The benchmark dose approach is a tool established in the 1980's that can be used to determine a POD for risk assessment (Crump 1984), and USEPA developed a draft technical document for its application (USEPA 2000). In principle, a BMD is a statistical approximation of a dose corresponding to a defined probability that a specified response will occur. For example a BMD_{10} (the effective mean dose necessary to produce a 10% response - referred to in this manuscript as BMD) and $BMDL_{10}$ (the lower 95th percent confidence interval of a 10% response on a dose-response curve - referred to in this manuscript as BMDL) are two common benchmark dose measures used as PODs for risk assessment. Recent studies with formaldehyde show that a benchmark dose approach can be used with transcript profiling data, hereafter referred to as toxicogenomics, to establish BMD/BMDL values based on dose-response evaluation at a genomic level (Andersen *et al.* 2008; Thomas *et al.* 2007; Yang *et al.* 2007).

Fenofibrate and methapyrilene are two prototypical nongenotoxic rodent carcinogens, one with a well established MOA (fenofibrate), and another carcinogen where the MOA is still being elucidated (methapyrilene). Fenofibrate is a rodent-specific nongenotoxic hepatocarcinogen and produces tumors in mice and rats following daily oral administration (Klaunig *et al.* 2003). There are several key events related to the

MOA of tumorigenesis, including activation of the PPARα receptor, proliferation of peroxisomes, expression of genes involved in lipid metabolism, oxidative stress resulting from excessive mitochondrial oxidation of fatty acids, with resulting DNA damage, and perturbation of the balance between cell proliferation and apoptosis ultimately leading to an increased formation of tumors (Klaunig *et al.* 2003; Lai 2004; Yu *et al.* 2003). Methapyrilene oral administration to rats resulted in hepatocellular carcinomas and cholangiocarcinomas (Lijinsky 1984; Lijinsky *et al.* 1980). With some exceptions (Althaus *et al.* 1982; Ashby *et al.* 1988; Turner *et al.* 1987) the majority of available data suggest that methapyrilene is not mutagenic (Iype *et al.* 1982; Lee *et al.* 1994; Mirsalis 1987; Oberly *et al.* 1993; Steinmetz *et al.* 1988). Although the MOA for methapyrilene tumorigenesis is not entirely clear, it is related to a sustained proliferative response as an adaptation to chronic periportal injury and cell death (Cunningham *et al.* 1995; Mercer *et al.* 2009).

In this study, these two compounds were tested to determine if toxicogenomic analyses could be used to establish response thresholds (BMD/BMDL) for PODs for cancer risk assessment and compared the results to more conventional methods. The high dimensional data set for fenofibrate and methapyrilene microarrays will be simplified to biological processes and pathways based on public ontologies. In the case of fenofibrate, the resulting biological processes should correspond with the established MOA for rodent carcinogenicity. Even though methapyrilene does not have an established MOA, BMD/BMDL values can be determined. Further, the identified toxicogenomic POD should be below tumorigenic doses and precursor events to carcinogenicity. A toxicogenomics approach to risk assessment for nongenotoxic carcinogens should provide a threshold for all potential key events necessary for the development of cancer.

Methods

Peroxisomal β-Oxidation (PBox) Assay

All reagents were obtained from Sigma Aldrich (St. Louis MO). Animals were administered fenofibrate (2 animals per dose group) by oral gavage at doses of 0 (vehicle control - 10% acacia), 1, 3, 10, 30, 100, 300, 600, and 1,000 mg/kg over four days. Liver samples were collected at necropsy, snap frozen in liquid nitrogen and stored at approximately 70°C for analysis. For protein assessment and peroxisomal β-oxidation (PBox), samples were homogenized in 250 mM Sucrose, 5 mM EDTA, and 20 mM Tris HCl buffer (pH 7.4) and centrifuged. Sample protein content was determined using the Coomassie® Plus Protein Assay (modification of the Bradford methodology) using bovine serum albumin (BSA). Individual sample homogenates were combined with an aliquot of Triton X-100 (1% v/v final), vortexed, centrifuged, and the resulting supernatant was placed on ice. PBox reaction mixture was prepared on the day of assay and was comprised of 0.05 mM HEPES (pH 7.4), 20 mM NAD, 330 mM DTT, 10 mM CoA, 1 mM FAD, 15 mg/mL BSA and 100 mM KCN. Sample supernatant and PBox reaction mixture were added to a Spectramax microtiter plate (molecular Devices, Sunnyvale CA) and the reaction was initiated by the addition of palmitoyl CoA to each well. The microtiter plate was then placed into a plate reader (set at 37°C) and the optical density (OD) was read at 340 nm. The rate of the reaction was calculated from the steepest linear portion of the reaction curve using an extinction coefficient of 6.22 x 10^3 M^{-1} cm^{-1} for NADH.

Animal Studies

Fenofibrate or vehicle (1% carboxymethylcellulose sodium, 0.5% sodium lauryl sulfate, 0.085% Povidone [Sigma, St. Louis, MO]) were delivered to female rats (n=3) in two daily oral gavage doses of 30 or 1000 mg/kg/day. Samples of the left lateral lobe of the liver (~100 mg) were collected in RNAlater™ (Ambion, Austin, TX) 4 hours after the

second gavage dose. Total RNA was extracted from the liver section homogenized in RNA STAT-60 (Tel-Test, Friendswood, TX) and purified with RNeasy columns (Qiagen, Ca). RNA preparation, sample hybridization, chip washing, and chip scanning were performed as described in the Affymetrix GeneChip® Expression Analysis Technical Manual (Revision 1) (http://www.affymetrix.com/index.affx). For methapyrilene, a dataset collected by a multiorganizational consortium sponsored by the Hepatotoxicity Working Group of the International Life Sciences Institute, Health and Environmental Sciences Institute (ILSI HESI) was used (Waring *et al.* 2004). The oral gavage doses in the study for methapyrilene were the vehicle control, 10 mg/kg/day and 100 mg/kg/day over a period of 1, 3, and 7 days. Further details of the experimental procedure were described in Waring *et al.* 2004. Raw methapyrilene microarray data was downloaded from ArrayExpress (http://www.ebi.ac.uk/microarray-as/ae/) (Brazma *et al.* 2003).

Microarray Analysis

RNA extracted from liver samples treated with fenofibrate and methapyrilene were analyzed using RGU34A and the MAS5.0 algorithm; chips were checked for quality control statistics using Gene Expression Counsel. Only the samples that were within bounds of the quality control statistics were used. Signal data was normalized based on a $\log_2$ (log base 2) transformation consistent with prior methodologies (Andersen *et al.* 2008; Yu *et al.* 2006). Following data transformation, the large dimensionality (8,799 probe sets total) of data was reduced using one way ANOVA $p \leq 0.05$ as a significant change and the Benjamini and Hochberg method for multiple testing correction using BMDExpress (v1.3) (Benjamini and Hochberg 1995; Yang *et al.* 2007). Heatmap visualization of the significant genes were generated using the TIGR Multiexperiment Viewer (Saeed *et al.* 2006; Saeed *et al.* 2003). This Heatmap included hierarchical clustering and Euclidean distance to measure differences between responses. The $\log_2$ ratio change over background was used to construct the Heatmap.

Benchmark Dose Calculations

The benchmark dose analysis was performed on the significantly perturbed genes using the BMDExpress (v1.3), which generates a BMD and BMDL (Yang *et al.* 2007). Since each study had only three doses (including vehicles), only the 1° and 2° (polynomial) curve fitting models were applied and the best fitting model was selected based on the nested-chi squared test. The benchmark dose analysis for the PBox assay was performed using BMDS (v2.0) available online (http://www.epa.gov/ncea/bmds/dwnldu.html). The output from the best fitting continuous model was reported following draft guidance from the USEPA (USEPA 2000).

Gene Ontology and Kegg Pathway Analyses

Two ontologies were analyzed to generate biological processes and pathway information from significantly changes genes: Gene Ontology (GO) (http://www.ebi.ac.uk/GOA/) and Kegg Pathway (http://www.genome.jp/kegg/pathway.html) (Ashburner *et al.* 2000; Kanehisa and Goto 2000; Kanehisa *et al.* 2006). The analysis was performed using DAVID (Database for Annotation, Visualization, and Integrated Discovery), which generates an EASE score or modified Fisher's Exact Test for each term; a cutoff of $p \leq 0.1$ was considered significant (Dennis *et al.* 2003; Huang da *et al.* 2009). The GO analysis focused mainly on biological processes as it seemed the most relevant for evaluation of key events. The highest hierarchical level (DAVID category - GOTERM_BP_1) was chosen to facilitate data reduction. This level contains all genes that exist in lower hierarchical levels; thus, the risk assessment based on BMD/BMDL calculations would be the most comprehensive. For each significant Kegg Pathway and GO term, a median BMD and BMDL and lower 95th percentile BMD and BMDL were calculated.

Results

Fenofibrate

Prior to toxicogenomic analysis, a dose-response study was conducted to identify a low and high dose for further evaluation (Figure 1). PBox is a biochemical measure of a biological key event (peroxisome proliferation) for liver carcinogenicity in rodents (Klaunig *et al.* 2003) and a precursor event that is necessary for the development of cancer. Doses greater than 10 mg/kg produced a significant increase (13-665%) in PBox activity over background. Cancer risk assessments are often based on rodent carcinogenicity studies that employ a low and high dose design, along with controls. To approximate this design, we chose the 30 and 1,000 mg/kg doses for microarray analysis of liver transcript profiles.

Fenofibrate dosing significantly altered signals for 41 probe sets out of 8,799 over the two doses (Table 1). A visual inspection of these genes showed that most were involved with activation of fatty acid metabolism and peroxisome proliferation consistent with the putative mechanism of fenofibrate. Hierarchical clustering of the individual animal data showed dose-dependent increases in expression of many of the genes in the latter categories; the highest fold-increase was seen with mitochondrial acyl-CoA thioesterase 1 (Figure 2). Since an increase or decrease in a single gene may not constitute a key event, we calculated the BMD and BMDL values for each significantly changed probe set as a measure of variability (Figure 3). While there were a large range of BMD values, the majority of the genes had BMDs (~70%) or BMDLs (~80%) less than 10 mg/kg/day. Interestingly, only one of the 4 genes with the highest BMD value is linked to the presumed MOA for fenofibrate (increase peroxisomal and mitochondrial fatty acid metabolism), carnitine palmitoyltransferase 1B; for the other three (MAXdimerization protein, HSP 27 and cell-death induced fragmentation factor) the linkage is less obvious. These genes may be a result of cell injury due to off-target

pharmacology at higher dose levels or effects observed (e.g. stress response) at a higher level of PPARα activation.

Two different approaches were used to reduce the dimensionality: Kegg and GO analyses. The pathways identified by these analyses corresponded well with the established MOA for fenofibrate. Kegg analysis identified those biochemical pathways significantly perturbed based on the number of genes in the pathway with a significant change in expression (Table 2). The pathways most affected following fenofibrate exposure were fatty acid metabolism; PPAR signaling pathway; and valine, leucine and isoleucine degradation. The highly significant GO term impacted was also related to metabolic processes (Table 3). There was a tight range for calculated BMD and BMDL values using Kegg and GO analyses. The range for the lower 95th percentile BMDLs, which are the most relevant values to determine a POD, was 2-4 mg/kg/day.

Methapyrilene

Having tested the approach with fenofibrate, we repeated the analysis using a dataset for methapyrilene (Waring *et al.* 2004), a well known rodent hepatocarcinogen, but with a less characterized MOA. The number of genes affected by methapyrilene was greater than fenofibrate and increased over time (Figure 4). After 1, 3, and 7 days of treatment, 210, 393, and 1,175 probe sets, respectively, were significantly changed ($p < 0.05$) over the two doses. The number of significantly perturbed genes increased from 1 to 3 days of treatment, and even further by 7 days.

Both Kegg and GO analyses were again used to reduce the complexity and determine BMD/BMDL values. The number of Kegg pathways significantly perturbed increased over time. Only one pathway, ribosomal protein production, was changed at all time points (Figure 5A). Other significant pathways following 7 days of exposure were related to metabolism, such as P450 metabolism, PPAR signaling pathway, arachidonic acid metabolism, cysteine metabolism, and androgen/estrogen metabolism.

GO analysis revealed a qualitatively similar pattern with biological processes changing over time (Figure 5B). The exception was genes related to metabolic process which was consistent at all time points, and this correlated well with the large degree of metabolic pathways perturbed in the Kegg analysis. Both Kegg and GO results are consistent with a model in which different biological processes and pathways, which likely include key events, are changed in 'waves' with different time-dependence. To determine if these differences across time points impacted the POD, the median and lower 95th percentile BMD and BMDL values were calculated for each time point for significant Kegg and GO terms. The BMD and BMDL values for Kegg (Figure 6A) and GO (Figure 6B) analyses were similar (within a factor of 2 or less for median and lower 95th percentile averages) at all time points. The BMD and BMDL values for Kegg (Table 4) and GO (Table 5) analyses following seven days of exposure, the time point where liver injury was observed (Waring *et al.* 2004), were within a tight range and the values were similar across Kegg and GO results. The range for the lower 95th percentile BMDLs was 7-53 mg/kg/day.

Discussion

Increasing public demand for cancer risk assessments for thousands of chemicals has important economic and health consequences (Foth and Hayes 2008; Williams *et al.* 2009). Elucidating mechanisms of carcinogenesis can inform these risk assessments, but is time and resource intensive. Recent advances using toxicogenomic analyses provide new opportunities for cancer risk assessment. Machine learning techniques, based on training set of compounds with known MOAs, can discriminate between genotoxic and nongenotoxic agents (Ellinger-Ziegelbauer *et al.* 2005; Fielden *et al.* 2007; Nie *et al.* 2006; Uehara *et al.* 2008). In addition, adaptation of standard dose-response methodology to toxicogenomic analyses allows thresholds to be set based on BMD/BMDL values for biological processes using well developed ontologies (Andersen

et al. 2008; Thomas *et al.* 2007). The latter approach is attractive since it does not require a training set of compounds and can be applied using accepted methodologies to define thresholds. This dose-response approach is also ideal for assessing risk for nongenotoxic carcinogens using a MOA framework based on current guidances (Boobis *et al.* 2006; USEPA 2005). Unfortunately, risk assessments for suspected nongenotoxic agents often default to a very conservative linear low dose approach since key events cannot be identified. To address this problem, methodology described by Anderson *et al.* (2008) was extended to derive BMD/BMDL values for two nongenotoxic carcinogens and establish PODs for cancer risk assessment.

Applying Systems Toxicology Approaches to a MOA Framework for Risk Assessment

The PODs for both fenofibrate and methapyrilene derived using both a systems approach and more standard endpoints are shown in Figure 7. In both cases, the systems approach resulted in a conservative threshold, with good agreement to those derived from standard endpoints. For example, with fenofibrate and methapyrilene, the lowest BMDL from Kegg and GO analyses (from the lower 95th percentile of both ontologies) are below the tumorigenic and precursor threshold estimated from actual carcinogenicity data. Although, it should be noted that these endpoints are subject to the same caveats applied to standard endpoints, e.g. they may be species specific and not be relevant for human carcinogenicity (Klaunig *et al.* 2003; Lampe and Kammerer 1990; Mirsalis 1987), the analysis illustrates the value of toxicogenomic analyses for dose-response assessment of nongenotoxic carcinogens. Additional validation is required, but this approach is informative for cancer risk assessment because limits of exposure for nongenotoxic carcinogens can be defined without a known MOA.

The toxicogenomic POD is conservative since the analysis includes all perturbed pathways. Nevertheless, this approach will result in a higher estimate compared to linear low dose extrapolation. Methapyrilene serves as a good example since it has a

cancer potency value (e.g. TD_{50}) established. For example, using the linear low dose approach (Sawyer *et al.* 1984) the exposure translating to a 1 in 1 million excess cancer risk for methapyrilene, calculated from the TD_{50} of 9 mg/kg/day (http://potency.berkeley.edu/), is a risk specific dose of 1 µg/day. However, the calculated acceptable daily intake (ADI) for methapyrilene using the toxicogenomic POD (7 mg/kg/day) and a 1000-fold uncertainty factor (10x for interindividual variability; 10x for interspecies variability; and 10x to adjust for chronic exposure - typical uncertainty factors used in risk assessment (USEPA 2002)), is 490 µg/day, i.e. 500-fold higher than that calculated from linear extrapolation. This difference in ADI is inherent to a threshold vs. a linear low dose extrapolation method and is not specific to the systems approach described herein.

<u>Strengths and Weakness of a Systems Approach</u>

This approach has a number of underlying assumptions that are worth discussing. The first important assumption in the method is that key event(s) linked to tumorigenesis are reflected by altered gene expression. This seems a reasonable assumption since it is hard to imagine a MOA for tumorigenesis that does not involve changes in gene expression. For example, oxidative stress, DNA damage, and increased cell proliferation, a few common carcinogenic MOAs, are all associated with changes in gene expression (Delker *et al.* 2006; Seidel *et al.* 2003; Seidel *et al.* 2006; Uehara *et al.* 2008). In addition, tissue injury, a process integral to many nongenotoxic carcinogens, is accompanied by change in gene expression and increases in the number of genes altered reflects the dose-response relationship for tissue injury caused by chemicals (Foster *et al.* 2007). Therefore, while a possibility, it seems unlikely that carcinogenicity would occur without gene expression.

Another assumption is that changes in clusters of genes identified by toxicogenomic analysis reflect the important key events. By reducing the complexity to

significantly perturbed pathways and biological processes, one could miss the impact of a single gene or an unknown pathway with a different dose-response behavior several ways: (1) the microarray itself may not contain a probe for an influential gene, (2) ontologies are based on our current understanding of biology and may not be fully described for all relevant pathways, and (3) limitations in the number of dosed groups may reduce statistical power and underestimate the number of significant gene responses. It seems unlikely that missing a single gene or pathway would result in a significant error in risk assessment. First, as shown by the fenofibrate analysis, the calculated BMD/BMDL values for individual genes had good agreement, with only a few outliers. Second, even if a single gene change or an unknown pathway constituted a key event, it seems very unlikely that tumors would form without significant alteration in other important biological processes, i.e. key events already captured in existing ontologies. This assumption is supported by the fact that the BMD/BMDL values for all significant Kegg and GO processes were within a tight range of values. It is also worth noting the USEPA Guidance does not require that all key events be used in the risk assessment, only key event(s) linked to a threshold-based MOA and which provide an adequate dose-response assessment (USEPA 2005). Thus, even if some genes or pathways were missed using a systems approach, our analysis suggests that this will not significantly influence the overall risk assessment. Moreover, similar uncertainties also apply to conventional risk assessments. As the technology improves (e.g. increased sensitivity with microarrays, or expanded ontologies) these limitations also will be reduced.

A third assumption underpinning the method is that the mechanism does not need to be completely elucidated to estimate cancer risk. Understanding the mechanism provides increased confidence in the risk assessment because it allows key events to be identified. However, this is often impractical since the mechanism of

tumorigenesis may be too complex to fully describe in a reasonable time frame. For example, methapyrilene was identified as a rodent carcinogen in the 1980's (Lijinsky *et al.* 1980), but to our knowledge the mechanism has not been established. This isn't surprising since our toxicogenomic analysis shows that the response is complex, with hundreds of genes and many pathways perturbed. Despite this complexity, our analysis supports the hypothesis that a POD can be established using BMD/BMDL values for individual pathways and these values were within a tight range. Therefore, while the complexity of the tumorigenesis makes it difficult to identify a key event, we suggest that this complex biology is reflected by changes in groups of genes that describe key events and that BMD/BMDL values based on these key events can be used in a MOA framework to assess the risk for a nongenotoxic carcinogen.

The final assumption is that the critical time window for a key event in carcinogenicity was not missed. The time course data for gene expression following methapyrilene dosing demonstrated that waves of unique biological processes and pathways occurred at different time points suggesting that measurement of a key event may depend on the time at which a measurement was made, increasing the complexity of the analysis (Figure 5). This creates a risk that without an identified MOA, the analysis will be performed at the wrong time point and the BMD/BMDL values will not accurately reflect risk. However, when we derived the BMD and BMDL values for methapyrilene at 1, 3, and 7 days, they did not differ significantly (Figure 6). Therefore, while changes in the key event(s) may differ across time points, the estimated threshold-dose from the toxicogenomic analysis remained the same. Admittedly, the time course used in these proof-of-concept studies was short relative to a carcinogenicity study. It is possible that the threshold may decrease over time if the carcinogen accumulates or if increased toxicity occurred when defense mechanisms are overwhelmed from

continuous exposure. An additional uncertainty factor may be advisable when using short-term toxicogenomic data for the risk assessment.

Conclusions

In conclusion, this work describes how a framework for carcinogenicity risk assessment can be developed using toxicogenomic data for nongenotoxic carcinogens, such as fenofibrate and methapyrilene. Data reduction to biological processes and pathways likely contain many of the key events associated with nongenotoxic carcinogenicity. BMD/BMDL values for biological processes and pathways are appropriate PODs used for cancer risk assessment as they represent thresholds for different key events. Studies with longer time points, and different species, tumor types, and nongenotoxic carcinogens could further validate and refine the framework. While conservative, this proposed risk assessment framework can provide a more practical and realistic risk estimate than defaulting to low dose linear extrapolation.

Tables

Table 1. Probe sets Significantly Perturbed Following Rat Oral Exposure to Fenofibrate (0, 30, and 1,000 mg/kg/day) for Two Days.

ID	Gene Name	Gene Symbol
J02749_at	ACETYL-COENZYME A ACYLTRANSFERASE 1	ACAA1
J02749_g_at	ACETYL-COENZYME A ACYLTRANSFERASE 1	ACAA1
X05341_at	ACETYL-COENZYME A ACYLTRANSFERASE 2 (MITOCHONDRIAL 3-OXOACYL-COENZYME A THIOLASE)	ACAA2
J02752_at	ACYL-COENZYME A OXIDASE 1, PALMITOYL	ACOX1
rc_AA799489_g_at	ACYL-COENZYME A OXIDASE 1, PALMITOYL	ACOX1
rc_AA893242_g_at	ACYL-COA SYNTHETASE LONG-CHAIN FAMILY MEMBER 1	ACSL1
rc_AI169758_at	APOLIPOPROTEIN C-III	APOC3
M17069_at	CALMODULIN 1	CALM3
rc_AA800243_at	CELL DEATH-INDUCING DNA FRAGMENTATION FACTOR, ALPHA SUBUNIT-LIKE EFFECTOR A (PREDICTED)	CIDEA_PREDICTED
D43623_at	CARNITINE PALMITOYLTRANSFERASE 1B	CPT1B
M26127_s_at	CYTOCHROME P450, FAMILY 1, SUBFAMILY A, POLYPEPTIDE 2	CYP1A2
X07259cds_s_at	CYTOCHROME P450, FAMILY 4, SUBFAMILY A, POLYPEPTIDE 22	CYP4A1
rc_AA924267_s_at	CYTOCHROME P450, FAMILY 4, SUBFAMILY A, POLYPEPTIDE 22	CYP4A1
rc_AI170568_s_at	DODECENOYL-COENZYME A DELTA ISOMERASE	DCI
D00729_g_at	DODECENOYL-COENZYME A DELTA ISOMERASE	DCI
D00569_g_at	2,4-DIENOYL COA REDUCTASE 1, MITOCHONDRIAL	DECR1
D00569_at	2,4-DIENOYL COA REDUCTASE 1, MITOCHONDRIAL	DECR1
U08976_at	ENOYL COENZYME A HYDRATASE 1, PEROXISOMAL	ECH1
K03249_at	ENOYL-COENZYME A, HYDRATASE/3-HYDROXYACYL COENZYME A DEHYDROGENASE	EHHADH
X60328_at	EPOXIDE HYDROLASE 2, CYTOPLASMIC	EPHX2
X05834_at	FIBRONECTIN 1	FN1
L00191cds#1_s_at	FIBRONECTIN 1	FN1
D16479_at	HYDROXYACYL-COENZYME A DEHYDROGENASE/3-KETOACYL-COENZYME A THIOLASE/ENOYL-COENZYME A HYDRATASE (TRIFUNCTIONAL PROTEIN), BETA SUBUNIT	HADHB
rc_AA998683_g_at	HEAT SHOCK 27KDA PROTEIN 1	HSPB1
rc_AI170613_g_at	HEAT SHOCK 10 KDA PROTEIN 1	HSPE1
rc_AA894332_at	INTEGRIN BETA 3	ITGB3
rc_AI171506_g_at	MALIC ENZYME 1	ME1
rc_AA891916_g_at	MEMBRANE INTERACTING PROTEIN OF RGS16	MIR16
Y09333_at	MITOCHONDRIAL ACYL-COA THIOESTERASE 1	MTE1

rc_AA892849_at	MAX DIMERIZATION PROTEIN 4 (PREDICTED)	MXD4_PREDICTED
X02918_at	PROLYL 4-HYDROXYLASE, BETA POLYPEPTIDE	P4HB
X02918_g_at	PROLYL 4-HYDROXYLASE, BETA POLYPEPTIDE	P4HB
AJ224120_at	PEROXISOMAL BIOGENESIS FACTOR 11A	PEX11A
M60103_at	PROTEIN TYROSINE PHOSPHATASE, RECEPTOR TYPE, F	PTPRF
rc_H31554_at	SIMILAR TO HYPOTHETICAL PROTEIN FLJ25416	RGD1559690_PREDICTED
rc_AI172293_at	STEROL-C4-METHYL OXIDASE-LIKE	SC4MOL
rc_AA893080_at	SELENOCYSTEINE LYASE	SCLY
rc_AA874999_at	SEC61 BETA SUBUNIT (PREDICTED)	SEC61B_PREDICTED
D00753_at	SERINE PROTEASE INHIBITOR	SERPINA3N
M24067_at	SERINE (OR CYSTEINE) PROTEINASE INHIBITOR, CLADE E, MEMBER 1	SERPINE1
J03621_at	SUCCINATE-COA LIGASE, GDP-FORMING, ALPHA SUBUNIT	SUCLG1

Table 2. Kegg Pathway Analysis Following Rat Oral Exposure to Fenofibrate for Two Days.

Kegg Term	Count[a]	Percent Relative to Total Terms[b]	P-Value[c]	BMD Median[d]	BMDL Median[e]	BMD lower 95th percentile[f]	BMDL lower 95th percentile[g]
Fatty acid metabolism	10	28.6%	4.09E-11	5.87	4.03	4.01	2.80
PPAR signaling pathway	9	25.7%	7.65E-08	5.87	4.03	4.67	3.24
Valine, leucine and isoleucine degradation	4	11.4%	2.54E-03	5.04	3.49	4.13	2.88
Bile acid biosynthesis	3	8.6%	2.02E-02	4.71	3.27	4.13	2.88
Arachidonic acid metabolism	3	8.6%	7.51E-02	5.87	4.03	5.14	3.55

a. Number of significantly perturbed genes (one way ANOVA $p \leq 0.05$ as a significant change and the Benjamini and Hochberg method for multiple testing correction) with a Kegg Term.

b. Percent of significantly perturbed genes (35) with Kegg Term. Note that one gene can be represented by multiple probe sets.

c. EASE score or modified Fisher's Exact Test.

d. Median benchmark dose (10% response over background - BMD) for all genes containing the Kegg term.

e. Median benchmark dose at the lower 95th percent confidence interval (BMDL) for all genes containing the Kegg term.

f. Lower 95th percentile of all BMD values containing a Kegg term.

g. Lower 95th percentile of all BMDL values containing a Kegg term.

Table 3. GO Analysis Following Oral Exposure to Fenofibrate for Two Days.

GO Term	Count [a]	Percent Relative to Total Terms [b]	P-Value [c]	BMD Median [d]	BMDL Median [e]	BMD Lower 95th percentile [f]	BMDL Lower 95th percentile [g]
Metabolic process	30	85.7%	1.66E-05	7.18	4.86	3.70	2.59
Response to stimulus	14	40.0%	5.42E-02	6.51	4.43	3.42	2.40

The GO terms analyzed using DAVID were biological processes of the highest hierarchical category (GOTERM_BP_1) to facilitate data reduction and identify key events.

a. Number of significantly perturbed genes (one way ANOVA $p \leq 0.05$ as a significant change and the Benjamini and Hochberg method for multiple testing correction) with a GO Term.

b. Percent of significantly perturbed genes (35) with the GO Term. Note that one gene can be represented by multiple probe sets.

c. EASE score or modified Fisher's Exact Test.

d. Median benchmark dose (10% response over background - BMD) for all genes containing the GO term.

e. Median benchmark dose at the lower 95th percent confidence interval (BMDL) for all genes containing the GO term.

f. Lower 95th percentile of all BMD values containing a GO term.

g. Lower 95th percentile of all BMDL values containing a GO term.

Table 4. Kegg Pathway Analysis Following Oral Exposure to Methapyrilene for 7 days.

GO Term	Count[a]	Percent Relative to Total Terms[b]	P-Value[c]	BMD Median[d]	BMDL Median[e]	BMD lower 95th Percentile[f]	BMDL lower 95th Percentile[g]
Ribosome	57	5.4%	2.53E-24	64.05	48.71	40.87	32.90
PPAR signaling pathway	30	2.8%	2.26E-04	73.93	54.80	18.58	9.51
Metabolism of xenobiotics by cytochrome P450	27	2.5%	1.40E-05	74.64	55.22	30.35	24.97
Arachidonic acid metabolism	24	2.3%	1.47E-04	77.58	56.94	18.95	9.57
Complement and coagulation cascades	23	2.2%	9.86E-04	72.17	53.74	40.64	32.74
Tryptophan metabolism	22	2.1%	1.44E-06	89.36	64.03	27.89	19.18
Androgen and estrogen metabolism	21	2.0%	3.78E-04	29.48	11.31	10.82	6.93
Fatty acid metabolism	21	2.0%	5.81E-04	91.38	64.70	18.95	9.57
Antigen processing and presentation	21	2.0%	7.14E-02	82.85	59.99	37.57	29.45
Glycolysis / Gluconeogenesis	17	1.6%	2.74E-02	83.83	60.54	38.35	29.36
Glycine, serine and threonine metabolism	15	1.4%	1.59E-03	54.37	42.38	17.98	9.01
Pyruvate metabolism	14	1.3%	2.18E-03	85.51	61.68	47.30	37.51
Valine, leucine and isoleucine degradation	13	1.2%	3.07E-02	91.95	65.00	49.56	39.10
Linoleic acid metabolism	13	1.2%	2.99E-03	68.98	51.79	33.89	27.70
Biosynthesis of steroids	12	1.1%	1.29E-03	91.02	64.74	26.81	11.06

Porphyrin and chlorophyll metabolism	12	1.1%	3.09E-02	77.67	57.01	41.55	32.21
Pyrimidine metabolism	12	1.1%	7.08E-02	81.22	58.21	34.22	25.47
Alanine and aspartate metabolism	11	1.0%	3.08E-02	63.38	48.25	12.18	7.51
Carbon fixation	10	0.9%	7.78E-03	68.93	51.64	37.07	22.94
Propanoate metabolism	10	0.9%	2.99E-02	88.57	63.17	46.92	37.22
gamma-Hexachlorocyclohexane degradation	9	0.8%	5.88E-03	67.52	50.89	34.89	24.30
ABC transporters - General	9	0.8%	2.82E-02	62.56	47.60	31.54	20.48
Sulfur metabolism	8	0.8%	1.40E-03	28.81	11.15	20.45	9.73
Nitrogen metabolism	8	0.8%	5.87E-02	86.62	64.38	30.77	22.44
beta-Alanine metabolism	8	0.8%	3.97E-02	74.55	50.50	44.80	35.73
Lysine degradation	8	0.8%	5.87E-02	92.36	67.24	52.16	40.71
Caffeine metabolism	8	0.8%	3.64E-03	91.00	64.49	67.47	50.86
Caprolactam degradation	6	0.6%	5.05E-02	92.36	67.24	71.79	53.48
Terpenoid biosynthesis	4	0.4%	5.20E-02	87.57	62.63	45.34	36.12

a. Number of significantly perturbed genes (one way ANOVA $p \leq 0.05$ as a significant change and the Benjamini and Hochberg method for multiple testing correction) with a Kegg Term.

b. Percent of significantly perturbed genes (1066) with the Kegg Term. Note that one gene can be represented by multiple probe sets.

c. EASE score or modified Fisher's Exact Test.

d. Median benchmark dose (10% response over background - BMD) for all genes containing the Kegg term.

e. Median benchmark dose at the lower 95th percent confidence interval (BMDL) for all genes containing the Kegg term.

f. Lower 95th percentile of all BMD values containing a Kegg term.

g. Lower 95th percentile of all BMDL values containing a Kegg term

Table 5. GO Analysis Following Oral Exposure to Methapyrilene for 7 days.

GO Term	Count [a]	Percent Relative to Total Terms Impacted [b]	P-Value [c]	BMD Median [d]	BMDL Median [e]	BMD Lower 95th Percentile [f]	BMD Lower 95th Percentile [g]
Metabolic process	638	59.9%	2.09E-33	77.95	56.78	19.53	9.69
Biological regulation	312	29.3%	1.85E-04	76.33	56.03	18.72	9.51
Response to stimulus	265	24.9%	1.14E-03	75.61	54.84	18.77	9.55
Localization	246	23.1%	3.84E-07	77.61	56.97	20.93	10.06
Developmental process	241	22.6%	4.51E-06	75.93	55.99	15.64	8.81
Establishment of localization	217	20.4%	1.70E-06	77.61	56.98	20.93	10.06
Immune system process	72	6.8%	8.31E-04	75.75	55.88	20.11	9.92
Reproduction	46	4.3%	1.65E-02	71.28	53.21	10.82	6.93
Growth	36	3.4%	5.49E-04	62.29	46.83	20.21	9.78
Reproductive process	29	2.7%	1.13E-02	66.15	49.52	10.37	6.74
Multi-organism process	27	2.5%	4.37E-03	62.18	45.80	11.93	7.45

The GO terms analyzed using DAVID were biological processes of the highest hierarchical category (GOTERM_BP_1) to facilitate data reduction and identify key events.

a. Number of significantly perturbed genes (one way ANOVA $p \leq 0.05$ as a significant change and the Benjamini and Hochberg method for multiple testing correction) with a GO Term.

b. Percent of significantly perturbed genes (1066) with the GO Term. Note that one gene can be represented by multiple probe sets.

c. EASE score or modified Fisher's Exact Test.

d. Median benchmark dose (10% response over background - BMD) for all genes containing the GO term.

e. Median benchmark dose at the lower 95th percent confidence interval (BMDL) for all genes containing the GO term.

f. Lower 95th percentile of all BMD values containing a GO term.

g. Lower 95th percentile of all BMDL values containing a GO term.

Figures

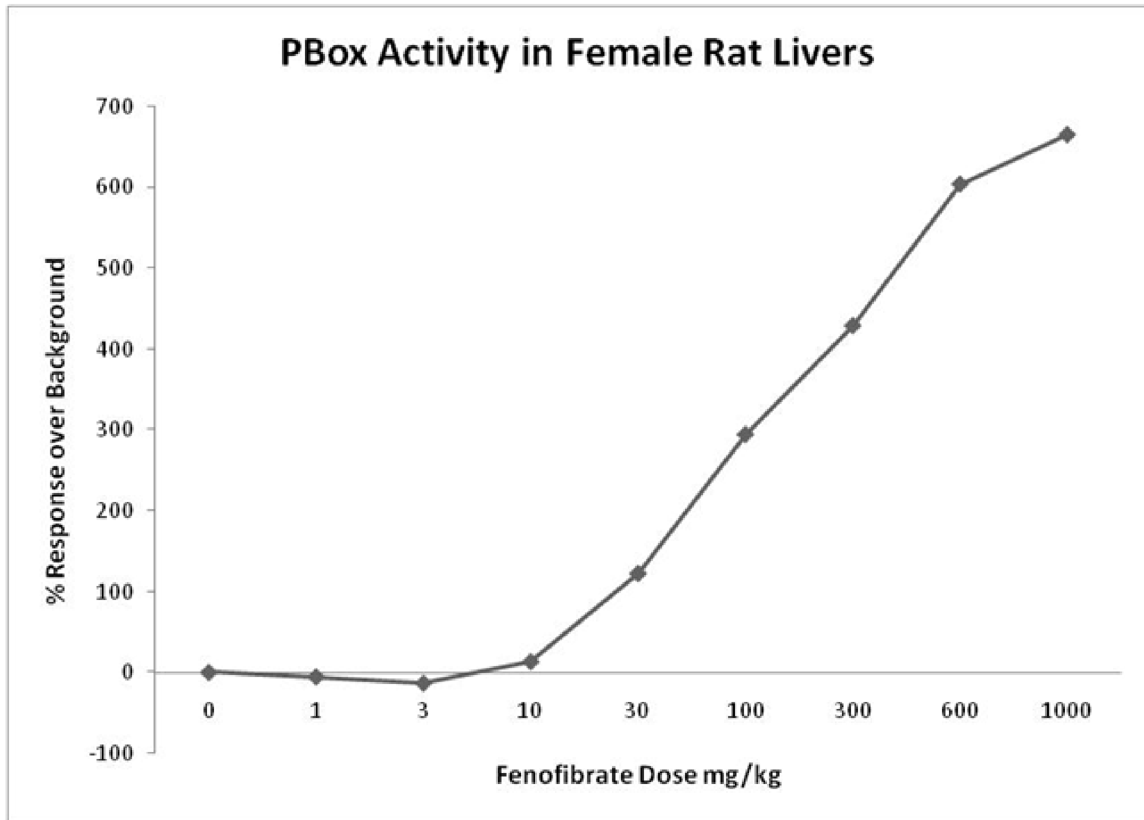


Figure 1. Peroxisomal β-oxidation (PBox) PBox activity in female rat livers following four days of oral fenofibrate. Each dose (n=2) is represented by a mean response relative to control at doses of 0-1000 mg/kg/day. Although response is represented as percent response over background, the units used to measure PBox activity was nmol NAD reduced / (min x mg protein).

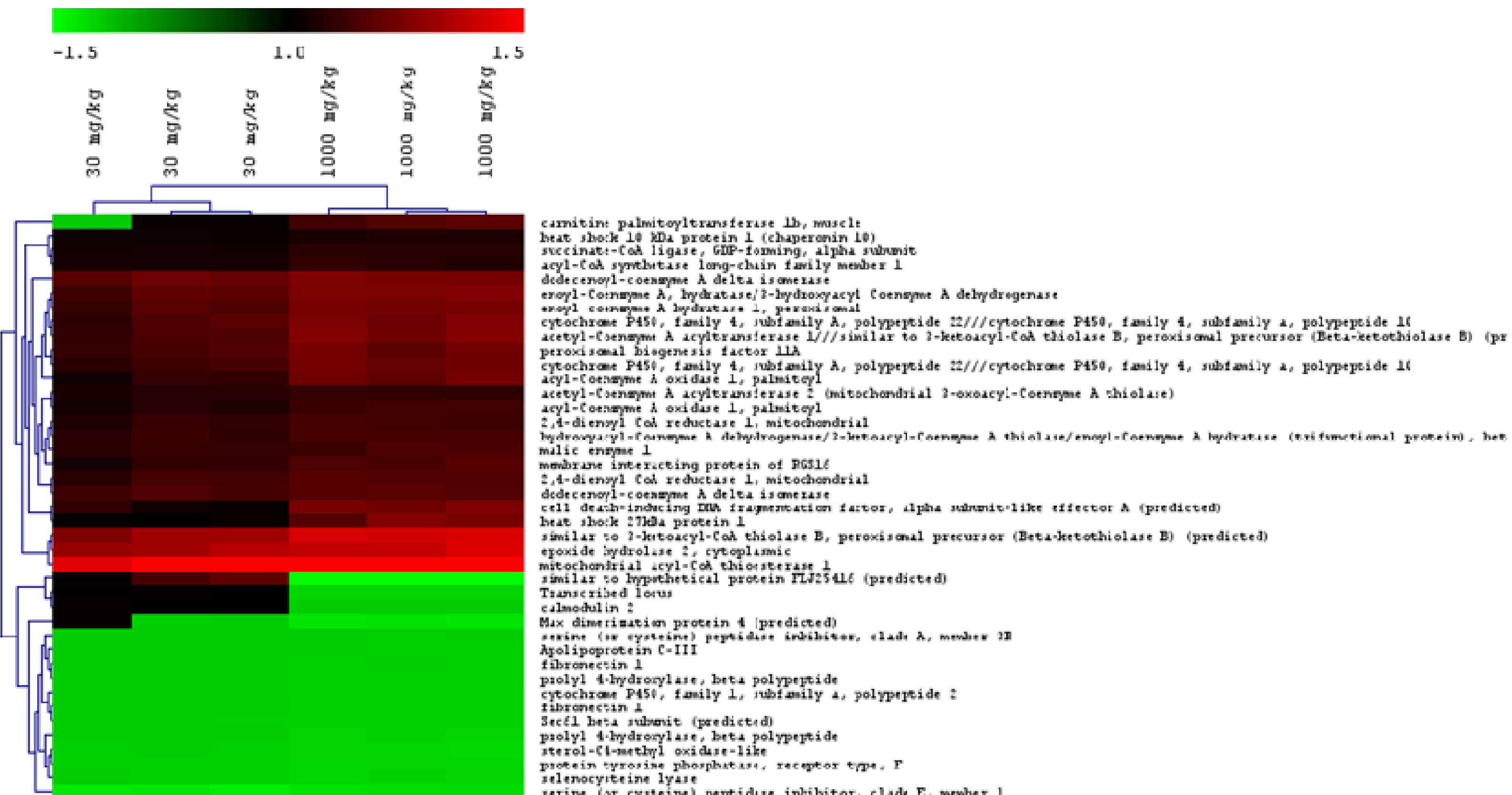


Figure 2. Hierarchical clustering of significantly perturbed genes in the liver following oral exposure to fenofibrate. Significantly perturbed genes were determined by using one way ANOVA $p \leq 0.05$ as a significant change and the Benjamini and Hochberg method for multiple testing correction. Colored responses represent a $\log_2$ signal relative to control.

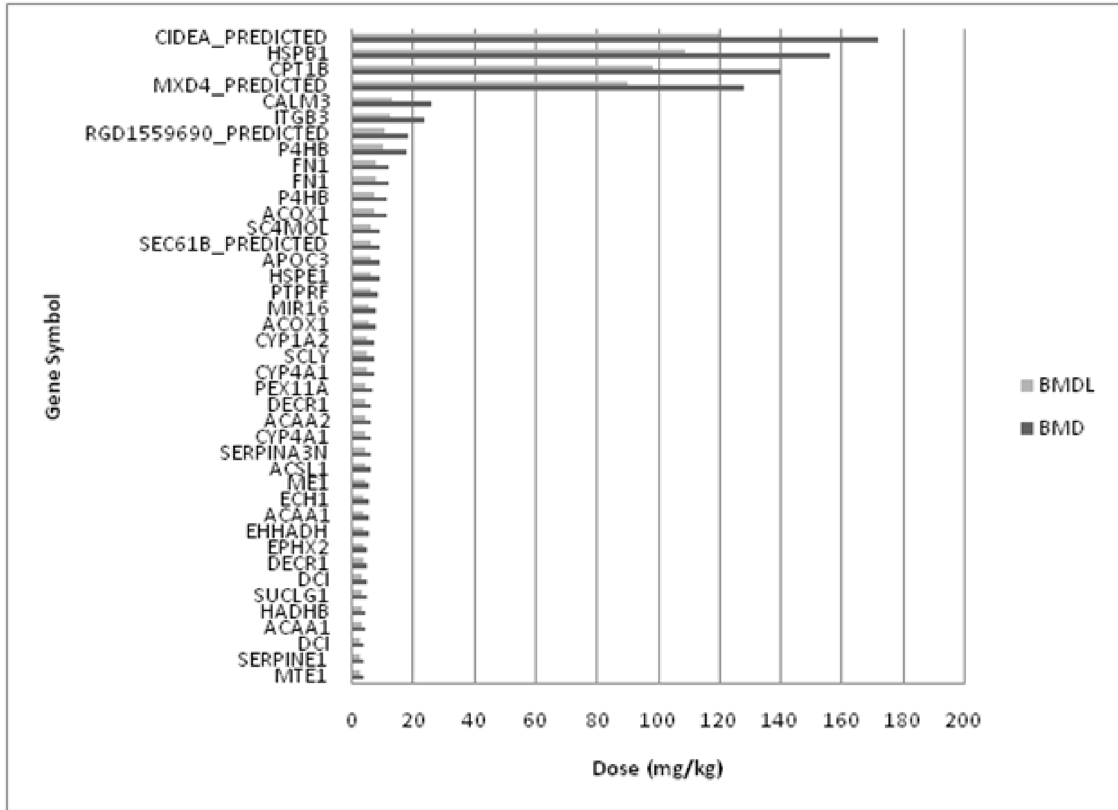


Figure 3. Benchmark dose analysis of each significantly perturbed gene following a two days exposure to fenofibrate at doses of 0, 30, and 1,000 mg/kg. BMD values represents a 10% gene response over background. BMDL represents a 10% response and the lower 95th percent confidence interval. Individual genes are represented by their symbolic notation.

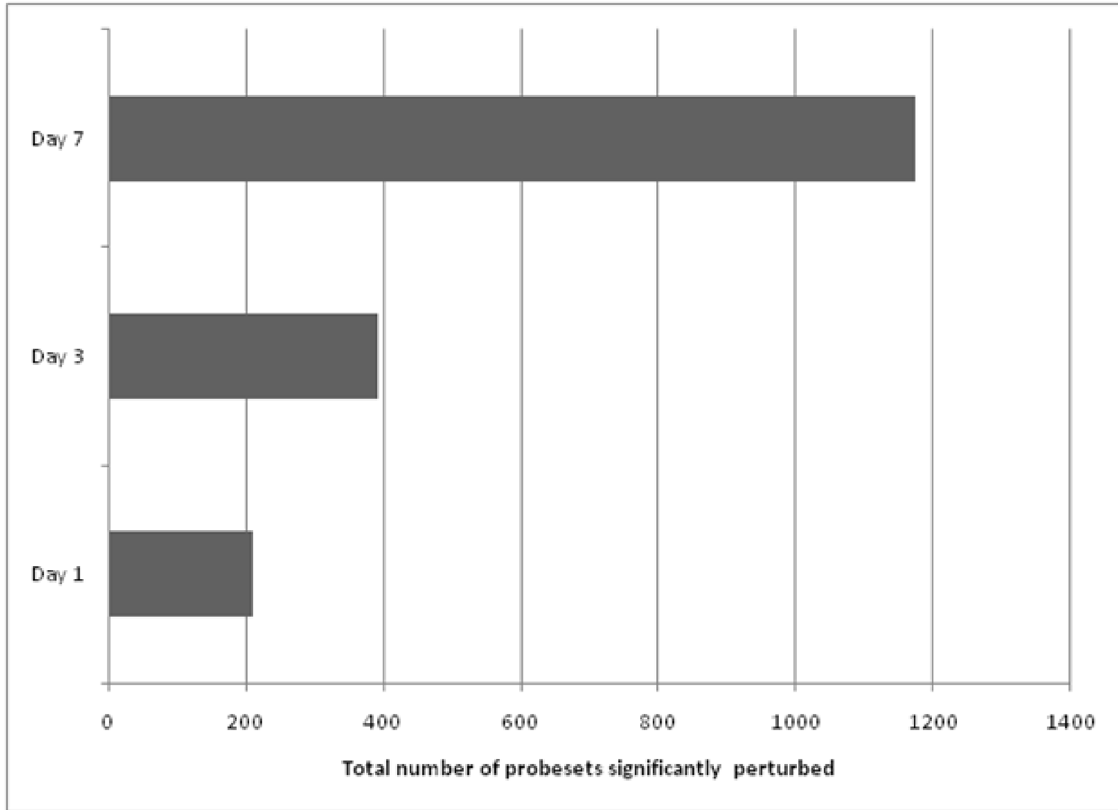


Figure 4. Time course for number of probe sets significantly perturbed following administration of methapyrilene. The oral gavage doses in the study for methapyrilene were the vehicle control, 10 mg/kg/day and 100 mg/kg/day over a period of 1, 3, and 7 days. Significantly perturbed genes were determined by using one way ANOVA $p \leq 0.05$ as a significant change and the Benjamini and Hochberg method for multiple testing correction.

A.

Percent of Terms Relative to Total

12%
10%
8%
6%
4%
2%
0%

1 Day
3 Days
7 Days

ABC transporters - General
Alanine and aspartate metabolism
Alkaloid biosynthesis II
Androgen and estrogen metabolism
Antigen processing and presentation
Arachidonic acid metabolism
beta-Alanine metabolism
Biosynthesis of steroids
Caffeine metabolism
Caprolactam degradation
Carbon fixation
Cell Communication
Complement and coagulation cascades
Fatty acid metabolism
gamma-Hexachlorocyclohexane degradation
Glycine, serine and threonine metabolism
Glycolysis / Gluconeogenesis
Linoleic acid metabolism
Lysine degradation
Metabolism of xenobiotics by cytochrome P450
Nitrogen metabolism
Porphyrin and chlorophyll metabolism
PPAR signaling pathway
Propanoate metabolism
Pyrimidine metabolism
Pyruvate metabolism
Ribosome
Sulfur metabolism
Terpenoid biosynthesis
Tryptophan metabolism
Tyrosine metabolism
Urea cycle and metabolism of amino groups
Valine, leucine and isoleucine degradation
C21-Steroid hormone metabolism

B.

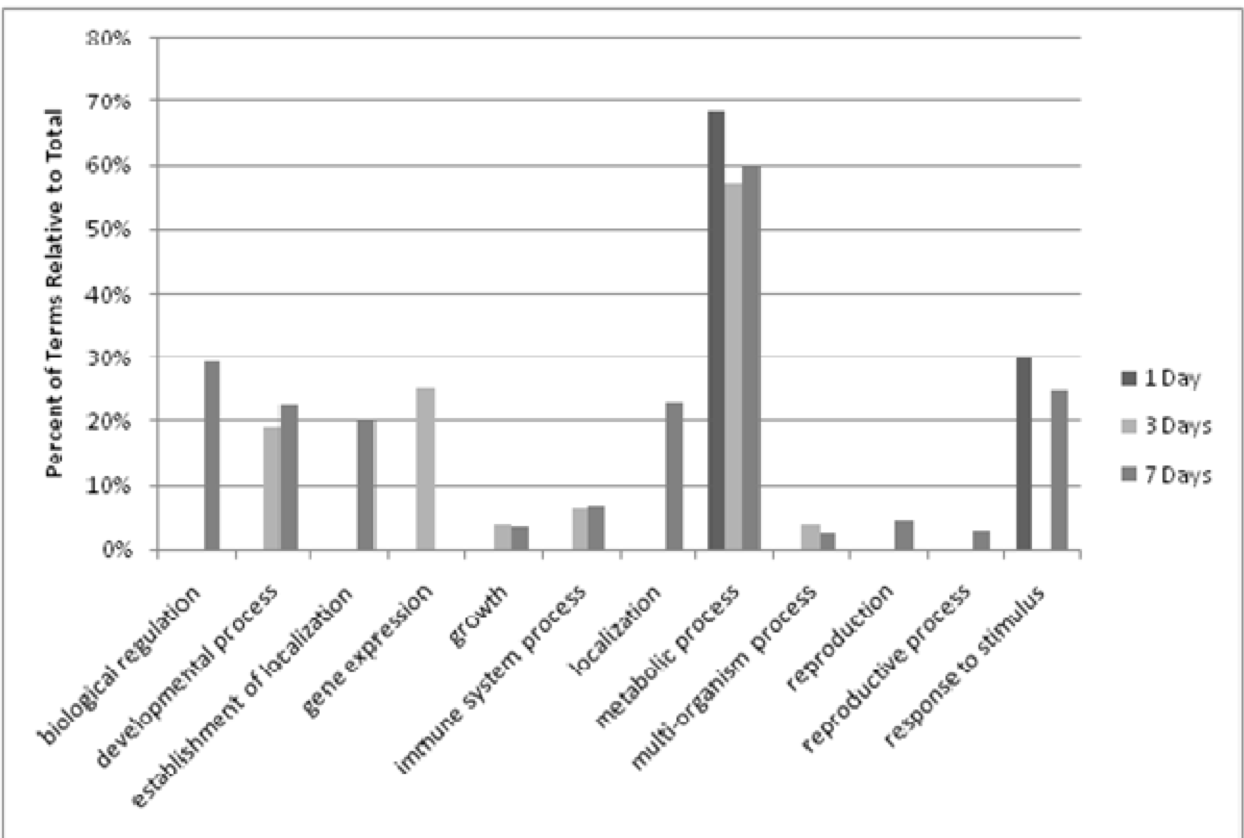
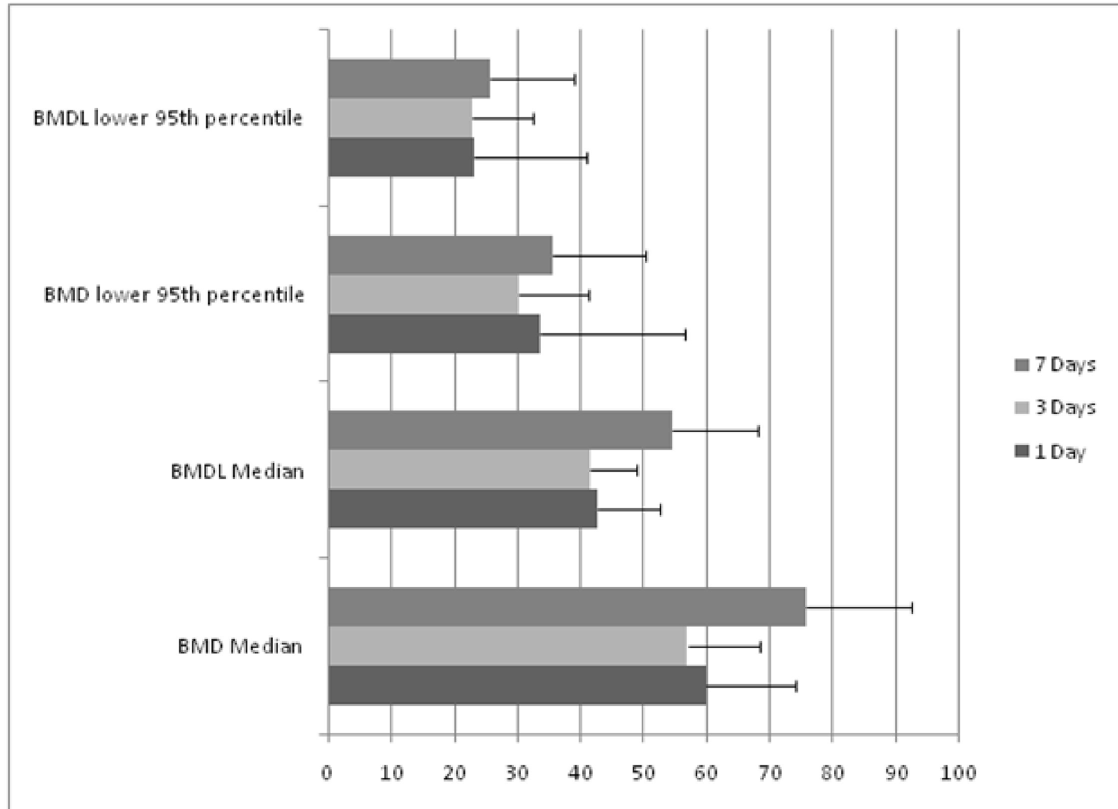


Figure 5. Methapyrilene (5A) Kegg and (5B) Gene Ontology (GO) time course analyses for significantly perturbed genes for all doses. Significant Kegg pathways or GO terms were determined by DAVID, which generates an EASE score or modified Fisher's Exact Test for each term, and a cutoff of $p \leq 0.1$ was applied to each Kegg and GO term. Reported are the percent of genes that contain the ontological term relative to the total number of significantly perturbed genes.

A.



B.

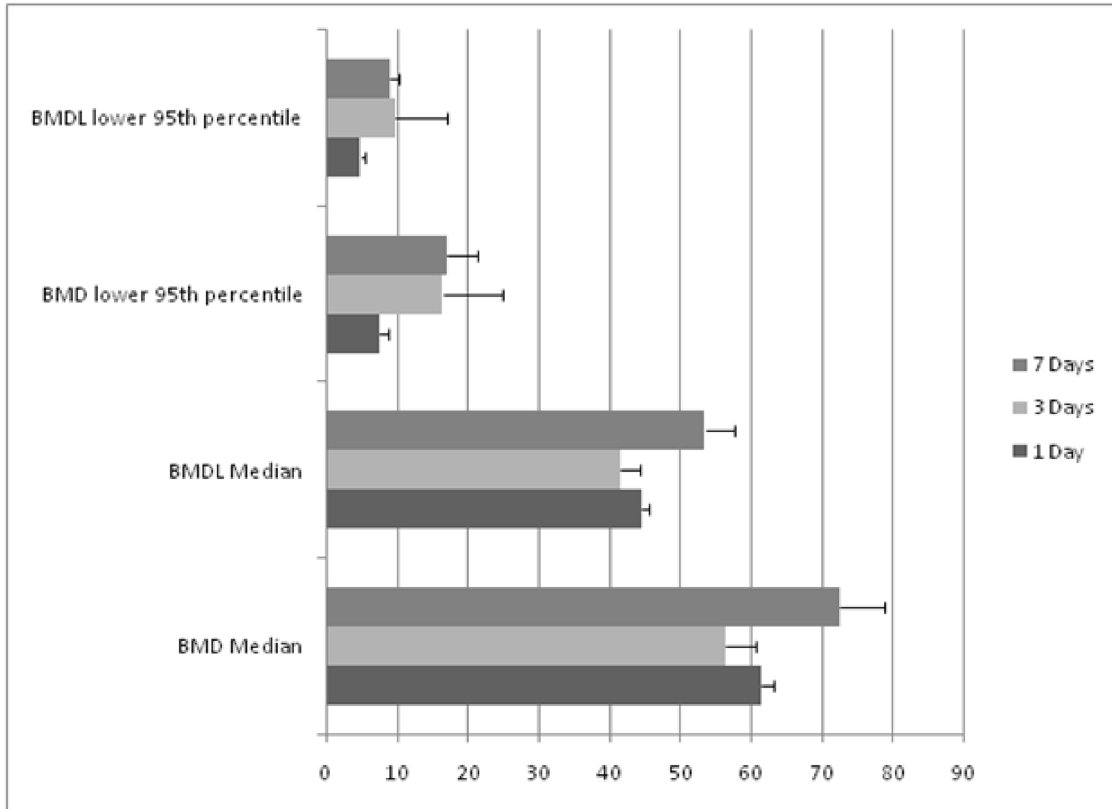


Figure 6. Mean and standard deviation BMD and BMDL values for (6A) Kegg Pathways and (6B) GO terms significantly impacted over 1, 3, and 7 days. Reported are the mean of the median values and the lower 95th percentile values for all significant Kegg pathway or GO terms.

A.

Tumorigenic dose in rodents – 200 mg/kg/day

PBox Activity
BMD – 6 mg/kg/day
BMDL – 5 mg/kg/day

Toxicogenomic lowest 95th percentile BMDL from GO and Kegg Analysis
2 mg/kg/day

B.

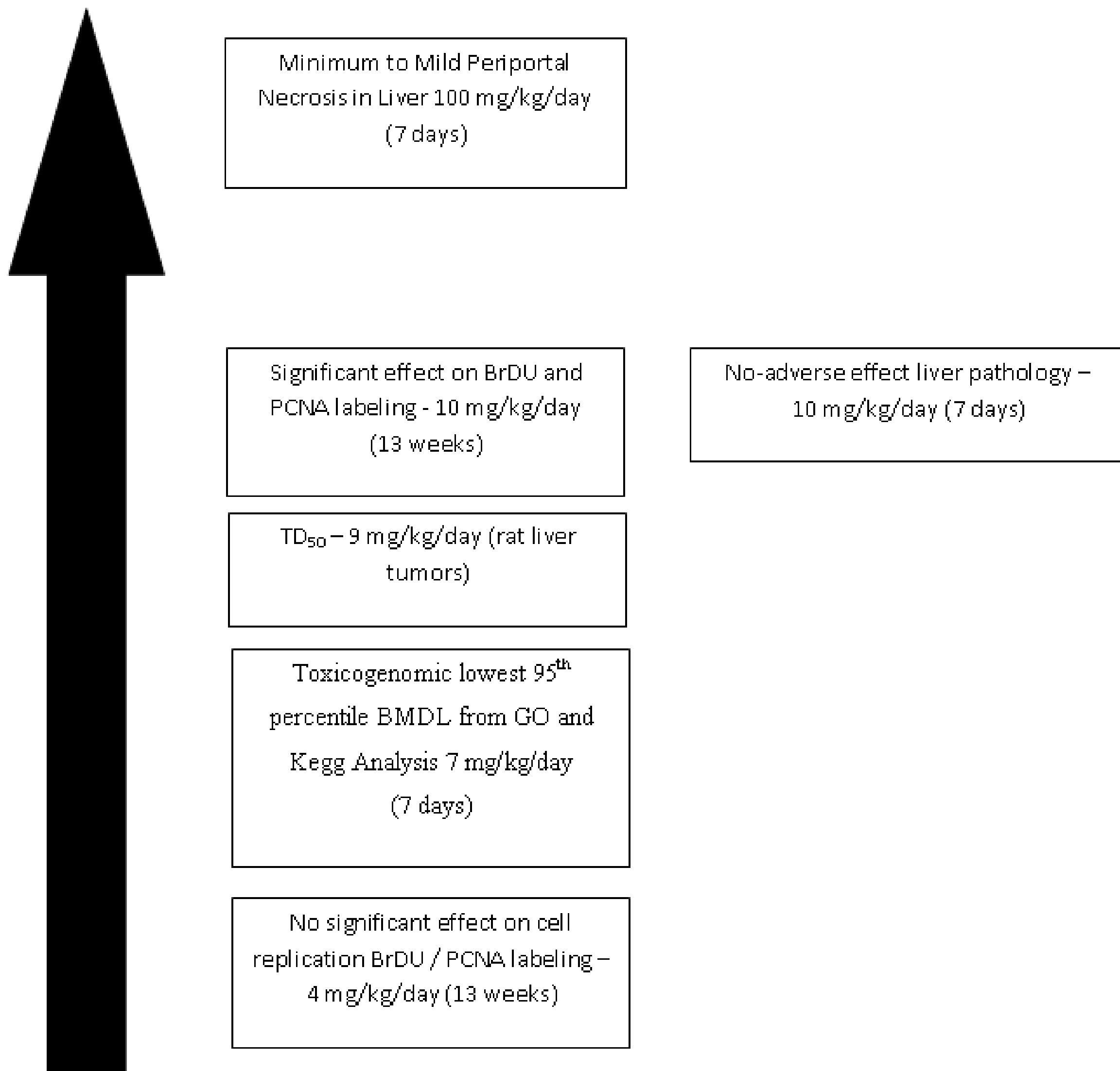


Figure 7. Comparison of toxicogenomics to other risk assessment endpoints for (7A) fenofibrate and (7B) methapyrilene for heptocarcinogenicity. PBox activity was considered the known key event identified for fenofibrate related to the MOA (PPARα activation) for carcinogenicity. The BMD and BMDL calculated for PBox activity was based on the best fitting model (Hill model). The tumorigenic dose for fenofibrate of was established from the package insert (http://dailymed.nlm.nih.gov/dailymed/about.cfm). While the MOA is unknown for methapyrilene, precursor effect-doses (liver injury and BrDU / PCNA labeling) (Waring *et al.* 2004; Cunningham *et al.* 1995) and tumorigenic dose (http://potency.berkeley.edu/) were identified from the literature.

DISCUSSION

Risk assessment is used to make decisions surrounding appropriate exposure levels of carcinogens in the environment. Oftentimes there is limited data to make a complete assessment. Therefore, the goal of these methodologies and three Specific Aims was to provide an assurance surrounding cancer risk, when limited data was available. Three critical needs were identified in cancer risk assessment: (1) mixtures of genotoxic compounds, (2) genotoxic metabolites, and (3) nongenotoxic carcinogens.

Advancement of Carcinogenicity Risk Assessment Framework

The results of the three aims are integral to the overall framework for carcinogenicity risk assessment. Figure 1 shows the decision tree for assessing cancer risk based on different types of information including techniques from the Specific Aims. The type of risk assessment framework that is required is dependent on the information available. Most compounds will not have carcinogenicity information due to the resources involved and high animal usage from cancer bioassays. Therefore genotoxicity is relied upon heavily to address carcinogenic potential. As addressed in Aim 1, a method was developed which can provide a more realistic assessment of carcinogenic risk than the conservative TTC default for genotoxic compounds without carcinogenicity information. A decision tree was designed so that non-potent compounds had a numerical cancer potency prediction and potent compounds used the TTC default. A similar approach may be applied for structurally similar compounds as dissimilar compounds, which uses separate risk assessments for each compound.

Animal carcinogenicity data does exist for many compounds, and should be used in concert with the genotoxicity test systems. Even though the compound may not be carcinogenic, the metabolite if genotoxic may still contain some carcinogenicity concerns. A new risk assessment model was developed in Aim 2 for genotoxic metabolites to assess their carcinogenic risk. This used exposure data in excretion

material in both animals and humans to determine the overall carcinogenic risk of a compound.

If a compound is determined to be carcinogenic in animals, one must determine the potential impact to humans. Genotoxic carcinogens are assumed to have no threshold and a linear low-dose extrapolation is typically applied. A threshold may exist for nongenotoxic carcinogens; therefore linear low-dose extrapolation may not be the most appropriate risk assessment methodology. If a threshold-based mode of action (MOA) is identified, then a less conservative risk assessment than linear low-dose extrapolation is applied. Aim 3 established methodology using toxicogenomics to analyze key events contained within the MOA. This analysis in combination with benchmark dose allows the identification of a threshold for a nongenotoxic carcinogen.

Each section of the decision matrix is critical to understanding the risk of a compound; thus the advancements in each area impacted the overall framework. The following sections are descriptions of how each Specific Aim advanced the overall framework for carcinogenicity risk assessment.

Specific Aim 1

Specific Aim 1 focused on providing a scientific analysis for mixtures of genotoxic compounds. Many compounds may either be positive in a mutagenicity assay or have a structure which is concerning from a carcinogenicity perspective, but no long-term carcinogenicity data exists in animals (Ashby and Paton 1993; Ashby and Tennant 1991; Zeiger *et al.* 1996). This is of particular concern for impurities in pharmaceuticals, where only limited toxicology data is available (Müller *et al.* 2006). Genotoxic impurities have a potential to cause a carcinogenic response in patients at levels where the genotoxicity / oncogenicity studies for the pharmaceutical substance could not theoretically detect the effect (Jacobson-Kram and Jacobs 2005; McGovern and Jacobson-Kram 2006). A contamination event of a genotoxic impurity, ethyl methanesulfonate, in the HIV drug

Viracept® underscored the importance of understanding the risk of genotoxic impurities in pharmaceuticals (Gerber and Toelle 2009; Muller and Singer 2009; Pozniak *et al.* 2009; Walker *et al.* 2009). Vital supplies of this life-saving medication were delayed until Roche was able to determine the safety impact of high levels of the genotoxin in their medication (CHMP 2008b). The threshold of toxicological concern (TTC) is an established default when genotoxicity but not carcinogenicity information is available for a chemical (Kroes *et al.* 2004; Müller *et al.* 2006). The TTC value was based on a distribution of known carcinogens, and a value was selected to represent a high probability of not exceeding a negligible excess cancer risk. The TTC has not only been used for pharmaceutical genotoxic impurities, but also for other industries such as food or personal care products (Blackburn *et al.* 2005; Felter *et al.* 2009; Kroes *et al.* 2004). Despite the utility of the TTC, it is a conservative default and improvements to the methodology can be made which provide a more accurate risk assessment. Methodology improvement for a single genotoxic compound could eventually be used for mixtures methodology.

The first study improved upon cancer risk assessment for a single genotoxic compound; we tested the hypothesis that carcinogenic potency can be predicted from a database of known carcinogens (cancer potency database - CPDB) (Chapter 2) (Gold *et al.* 1999; Gold *et al.* 2005; Gold *et al.* 1991). Models were successfully developed which predicted cancer potency in rats and mice. A framework was established from these models which: (1) predicts if a compound's carcinogenicity is potent or not potent, (2) makes a quantitative prediction of cancer risk if a prediction is considered not potent, and (3) defaults to the existing TTC for compounds considered potent or if indeterminate/unreliable predictions are made. These models errored to be conservative; thus, increased the number of false positives (compounds predicted to be potent that were not potent). Although conservative, this framework would be a

significant improvement to cancer risk assessment for a single genotoxic compound as realistic risk assessments can be made in certain instances in lieu of the more conservative TTC default. Once we improved the risk assessment for a single genotoxic compound, we focused on mixtures.

No methodology is yet available for multiple genotoxic compounds. Yet in reality we are exposed to a mixture of genotoxic compounds. This is was of particular concern for genotoxic impurities and reflected in the EMEA / FDA guidance documents (CHMP 2008a; Jacobson-Kram and Jacobs 2005; USFDA 2008b). Structurally similar compounds were to be treated with independent limits, while structurally similar compounds were to be treated like they were the same compound, thus grouping all the limits into one combined limit such that the total exposure would not exceed the TTC. The rationale was based on judgment and not on science. Therefore, our next study was to test the validity of current existing guidances.

The goal of our next study in Aim 1 was to develop a methodology to assess the risk of mixtures of genotoxic compounds (Chapter 3). Our hypothesis was that the excess cancer risk will be higher for structurally similar compounds than for structurally dissimilar compounds. First, we categorized all compounds from the CPDB by alert structure (structure that is correlated with carcinogenic activity), and if a compound did not belong in a category we considered it “other”. If structurally similar compounds behaved similarly from a biological perspective, then the carcinogenic risks in each category should correlate with each other. However, our results showed that the carcinogenic risks did not correlate which each other, which is not necessarily surprising. Structurally similar compounds are assumed to have a similar mechanism of action, which may imply risk values that are close to each other. However, slight changes in structure could influence the ability for a compound to be absorbed, metabolized and penetrate into the target organ. In fact, many of the compounds with the same alerting

structure still differed substantially with the rest of the moiety, or had an additional structural alert. Therefore, there was no evidence that compounds with the same alert structure would have correlated risks.

Despite evidence to the contrary, a second analysis was performed assuming structurally similar compounds would have correlated risks to determine if this would significantly impact the overall risk assessment. Monte Carlo simulations were performed for two scenarios: (1) structurally similar compounds with carcinogenic risks that were purely random, and (2) structurally dissimilar compounds with risks that were within ± 10-fold from the first compound. Despite the original hypothesis, total cancer risk decreased slightly for structurally similar compounds (correlated risks) relative to structurally unrelated compounds (random risks). Therefore, based on these two analyses, even if it were assumed that genotoxic compounds related in structure had similar carcinogenic potencies, it should not result in a lower TTC than for compounds structurally unrelated to each other.

In conclusion, Aim 1 was successfully completed by improving upon methodology for a single genotoxic compound and addressing cancer risk assessment for mixtures of genotoxic compounds. This work tested and expanded current methodology to improve upon current regulatory guidances. The major finding was that multiple structurally similar genotoxins would not likely provide a greater carcinogenic risk than structurally dissimilar genotoxins.

Specific Aim 2

Specific Aim 2 focused on the risk assessment of genotoxic metabolites. Metabolism is an essential part of understanding the safety of a consumed compound. Compounds when ingested undergo extensive biotransformation, yielding a variety of systemic metabolites. Safety testing for a compound requires an understanding of the potential toxicity of these metabolites (Baillie *et al.* 2002). Despite the need for

evaluating genetic safety of metabolites, no practical recommendations are available. A working group which convened at the 4th International Workshop of Genotoxicity Tests (IWGT) acknowledged the need for a practical strategy to respond to documented human metabolite exposures. They suggested that the TTC, similar to the value derived for pharmaceutical impurities described in Aim 1, may be used to support a risk assessment approach for genotoxic human metabolites (Ku *et al.* 2007). Therefore, the hypothesis was that the impurity-based TTC could be applied as a practical risk assessment tool for metabolites.

Prior to testing the hypothesis, the elements of risk assessment (e.g. hazard identification, exposure assessment, and risk characterization) were incorporated to determine a risk assessment framework for genotoxic metabolites (Chapter 4). For hazard identification, we found that it is less relevant to conduct a genotoxicity assessment for metabolites identified in the plasma. Genotoxic metabolites can have short half-lives in the bloodstream and difficult to measure. The best way to identify potential genotoxic hazards was to identify metabolites in the excretion material and then derive upstream metabolites. These "likely" metabolites can be analyzed based on structure or machine learning methods (e.g. *in silico*) to determine if in there is a potential for these compounds to have genotoxic activity (Ashby and Paton 1993; Hayashi *et al.* 2005; Snyder *et al.* 2004; White *et al.* 2003). Follow-up Ames testing if necessary can confirm predictions. Once concluded to be genotoxic, then it is possible to determine potential exposure to the metabolite.

Exposure to genotoxic metabolites must be quantified so it can be used in to assess risk. As with hazard identification, excretion material was determined to be a better estimate for exposure to genotoxic metabolites. Genotoxic substances can either concentrate in the tissues or bind to protein; thus, if measuring in the plasma one could underestimate the actual exposure of the metabolite in the body (Smith and Obach

2006). Excretion material can be used to calculate metabolite total body burden - total systemic exposure of the metabolite throughout the body (Smith and Obach 2005). The percent excretion, relative to dose, from downstream metabolites can then be used to determine the likely percent excretion of upstream metabolites. Once exposure was quantified in both animals and humans, it was possible to test if the TTC was appropriate for risk assessment.

Despite the hypothesis, it was found that the TTC used for pharmaceutical impurities was not adequate to derive risk for genotoxic metabolites. While advancements have been made to measure metabolites at low levels, it is unlikely that a metabolite can be measured at levels as low as the TTC developed for impurities. For example, assuming a 100 mg dose, the acceptable level of a genotoxic metabolite would be 0.0015%, which would be impossible to measure; whereas, it is possible to measure pharmaceutical impurities at extremely low (ppm) levels (Humfrey 2007; Pierson *et al.* 2009). Also, while it is possible to reduce impurity levels, it is not possible to reduce metabolite levels. Once a genotoxic metabolite is identified, it is then a part of the risk of a pharmaceutical that cannot be mitigated. This means there is a different level of risk tolerance where metabolites are part of the risk/benefit equation for a pharmaceutical; whereas, impurities are assumed to have no benefit and a lower level of risk is tolerated (Kasper 2004; McGovern and Jacobson-Kram 2006). For these reasons, we developed a model to derive a metabolite-based TTC.

Using the same CPDB database as the impurity-based TTC, we developed a new model for assessing the risk of genotoxic metabolites and used as a foundation for a metabolite TTC. The model incorporated exposure in humans and animals, and took into account the probability that a genotoxic compound is a carcinogen. The model for human exposure was developed to determine the probability that a carcinogen would be of negligible excess cancer risk defined as 1 incidence in 100,000 (10^{-5}). The model for

animal exposure was developed to determine the probability that a carcinogen would result in a tumorigenic response in animals, or in essence be "detected" in a bioassay. Finally, the probability that a genotoxic substance is a human carcinogen was fixed based on prior experience as 50% (Barlow *et al.* 2001). Each of these three probabilities were incorporated together to determine the total probability that a genotoxic metabolite would result in a 10^{-5} excess cancer risk. The next step was to use this model to generate a metabolite-based TTC.

Different animal and human dose combinations were tested in the model to determine some generalities and help generate a metabolite-based TTC. There were some generalities that emerged from this exercise. These generalities indicated that there was a high probability of not exceeding a 10^{-5} excess cancer risk under certain scenarios: (1) if an animal metabolite dose was ≥ 10 mg/kg/day, (2) if the margin was ≥1000x when dividing animal exposure by human exposure, or (3) if human exposure ≤ 1.5 µg/day, which confirmed the observations used to derive the impurity-based TTC (Kroes *et al.* 2004). These generalities in combination with the risk framework were adequate measures of a metabolite-based TTC.

In conclusion, while the impurity-based TTC was not found to be an effective methodology for metabolites, an alternative metabolite-based TTC was derived. Certain generalities emerged from model development that were used to determine the probability of not exceeding a 10^{-5} excess cancer risk. These generalities and the model can be used to determine the safety of an identified genotoxic metabolite and its utility was established using case studies as described in Chapter 4.

Specific Aim 3

In Aim 3, the focus was to establish a risk assessment approach for nongenotoxic carcinogens. The difference between genotoxic and nongenotoxic carcinogens is that genotoxic carcinogens cause direct damage to DNA resulting in

carcinogenicity (Barlow *et al.* 2006). The mechanism of nongenotoxic carcinogenicity can be through a variety of mechanisms and DNA damage can occur but through an indirect mechanism (Butterworth 2006; USEPA 2005a). The assumption in the prior two Aims was that genotoxic compounds or metabolites had no threshold. Nongenotoxic carcinogens may act via a threshold, which is the dose-cutoff where exposure below the dose should not have a carcinogenic effect (Alden 2000; USEPA 2005a). A threshold can be established by understanding the MOA or a description of key event(s) necessary to result in carcinogenicity (USEPA 2005a). A MOA is different from mechanism, which requires a detailed causal understanding of tumorigenesis many times at the molecular level. However, it can be difficult to establish a MOA since it can be experimentally time-consuming and there can be multiple MOAs with multiple key events. Therefore, the goal of Aim 3 was to improve upon existing carcinogenicity methods by identifying a threshold for key events of nongenotoxic carcinogens, otherwise known as the point of departure (POD). Similar to the TTC, this can be used to determine a safe level for a nongenotoxic carcinogen.

The hypothesis tested in Chapter 5 was that the POD derived from genomic analysis of key events could be determined and it was below tumorigenic and precursor events. This hypothesis was tested with two known nongenotoxic carcinogens, fenofibrate and methapyrilene, using toxicogenomics. Administration of each compound resulted in hepatocarcinogenicity in long-term oncogenicity studies, which was not associated with direct damage to the DNA (Lijinsky 1984; Lijinsky and Kovatch 1986; Lijinsky *et al.* 1992; Lijinsky *et al.* 1980; Mirsalis 1987; PI 2006). The MOA has been established for fenofibrate while the MOA has not been established for methapyrilene. Fenofibrate activates the PPARα receptor which results in peroxisomal proliferation and at a certain threshold this overwhelms the organism's liver homeostasis resulting in heptocarcinogenicity when administered over a chronic duration (Gonzalez *et al.* 1998;

Klaunig *et al.* 2003; Yu *et al.* 2003). Methapyrilene is a nongenotoxic carcinogen where a MOA has not yet been established, but it is related to a sustained proliferative response as an adaptation to chronic periportal injury and cell death (Cunningham *et al.* 1995; Mercer *et al.* 2009). While these nongenotoxic carcinogens may be species specific (Klaunig *et al.* 2003; Mirsalis 1987), they serve as adequate examples to test the hypothesis.

Ontological analysis of microarray data was determined to be effective for identification of key events for nongenotoxic carcinogenicity. The basic assumption is that gene expression is impacted by nongenotoxic carcinogenicity, which is reasonable (Chapter 5). Well-developed ontologies such as Kegg Pathways (Kanehisa and Goto 2000; Kanehisa *et al.* 2006) and Gene Ontology (GO) (Ashburner *et al.* 2000) were analyzed (Dennis *et al.* 2003) to identify the biological processes and pathways that are impacted by perturbed genes. It is also reasonable to assume that these biological processes and pathways contain key events for carcinogenicity (Chapter 5). While it may be impossible to know if all key events were analyzed, identifying key event(s) related to dose-response activity is consistent with current regulatory guidances (USEPA 2005a). Adaptation of standard dose-response methodology, e.g. benchmark dose, to genomics allowed thresholds to be set based on these ontologies; thus, this led to thresholds for key events (Andersen *et al.* 2008; Thomas *et al.* 2007; Yang *et al.* 2007). Therefore, toxicogenomics filled a critical gap for cancer risk assessment of nongenotoxic carcinogens by providing a dose-response for potential key events; this analysis served as a POD for cancer risk assessment in lieu of conservative low dose linear extrapolation.

A comparison of the toxicogenomic POD with tumorigenic or precursor doses confirmed the hypothesis; the toxicogenomic PODs were below tumorigenic and precursor effects. The toxicogenomic POD of 2 mg/kg/day was slightly below the

threshold for PBox activity (6-7 mg/kg/day) and significantly below the tumorigenic dose (200 mg/kg/day) (PI 2006). Since no single MOA was known for methapyrilene carcinogenicity, several biological markers were captured from the literature for validation. The toxicogenomic POD (7 mg/kg/day) was slightly below the subchronic mitogenic dose (10 mg/kg/day) (Cunningham 1996; Cunningham *et al.* 1995) and the tumorigenic dose (TD_{50} = 9 mg/kg/day) (http://potency.berkeley.edu/chempages/METHAPYRILENE.HCl.html) and well below the acute (7 days) dose which resulted in liver periportal necrosis (100 mg/kg/day) (Waring *et al.* 2004). Therefore, the toxicogenomic POD was validated in these two examples to be below other effect-doses for carcinogenicity.

In conclusion, a threshold for nongenotoxic carcinogens could be determined via an analysis of key events and dose-response using toxicogenomics. The approach was validated with two nongenotoxic carcinogens, where their toxigenomic PODs were below tumorigenic or precursor effect-doses. Further study would validate the method: longer dosing regimens; and different species, tumor sites, and nongenotoxic compounds. The overall impact of this study is that it provides a framework to determine the safe dose of a nongenotoxic carcinogen without the time and resources necessary using traditional methodology.

Conclusions

This dissertation successfully tested and demonstrated a risk assessment framework for mixtures of genotoxic compounds, genotoxic metabolites, and nongenotoxic carcinogens, when data was limited. While the overall goal for protection of public health was the same, each type of risk assessment required a unique approach. An informatics approach was critical since it was necessary to retrieve, annotate and analyze information from existing databases and ontologies. Although statistical analysis was important, it was equally important to understand the structure of

the data, the relationship among organizing themes within the data, e.g. gene ontologies, and how these related to biologically relevant conclusions. Estalishing the application of informatics technology at this interface between biological and statistical analyses is essential since the demand for the safety assessment of chemicals continues to grow with pressures of reduced animal testing. As shown by the risk assessment framework addressed by the Specific Aims (Figure 1), informatics will become an increasingly important part of risk assessment. Future research will be needed to improve upon existing methodology and examples of successful implementation to show its practical use.

Figures

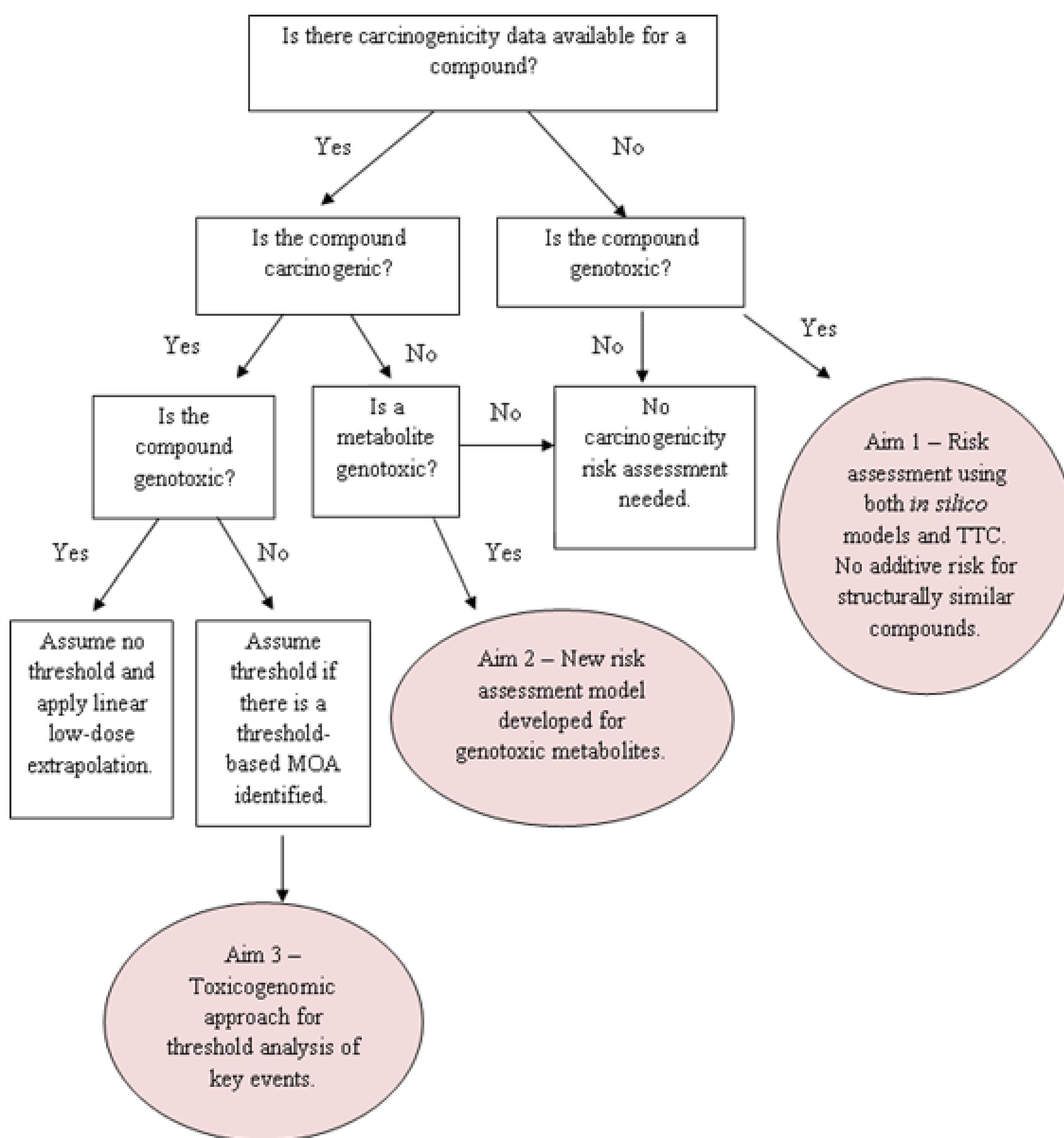


Figure 1. Advancement of the overall framework for carcinogenicity risk assessment from the three Specific Aims.

REFERENCES

Alden, C. L. (2000). Safety assessment for non-genotoxic rodent carcinogens: curves, low-dose extrapolations, and mechanisms in carcinogenesis. Hum. Exp. Toxicol. 19, 557-560.

Althaus, F. R., Lawrence, S. D., Sattler, G. L., and Pitot, H. C. (1982). DNA damage induced by the antihistaminic drug methapyrilene hydrochloride. Mutat. Res. 103, 213-218.

Ames, B. N., Shigenaga, M. K., and Gold, L. S. (1993). DNA lesions, inducible DNA repair, and cell division: three key factors in mutagenesis and carcinogenesis. Environ. Health Perspect. 101, 35-44.

Andersen, M. E., Clewell, H. J., 3rd, Bermudez, E., Willson, G. A., and Thomas, R. S. (2008). Genomic signatures and dose-dependent transitions in nasal epithelial responses to inhaled formaldehyde in the rat. Toxicol. Sci. 105, 368-383.

Anderson, L. M. (2004). Introduction and overview. Perinatal carcinogenesis: growing a node for epidemiology, risk management, and animal studies. Toxicol. Appl. Pharmacol. 199, 85-90.

Argentine, M. D., Owens, P. K., and Olsen, B. A. (2007). Strategies for the investigation and control of process-related impurities in drug substances. Adv. Drug Deliv. Rev. 59, 12-28.

Arni, P., and Hertner, T. (1997). Chromosomal aberrations in vitro induced by aneugens. Mutat. Res. 379, 83-93.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., and Sherlock, G. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. 25, 25-29.

Ashby, J., and Paton, D. (1993). The influence of chemical structure on the extent and sites of carcinogenesis for 522 rodent carcinogens and 55 different human carcinogen exposures. Mutat. Res. 286, 3-74.

Ashby, J., and Tennant, R. W. (1991). Definitive relationships among chemical structure, carcinogenicity and mutagenicity for 301 chemicals tested by the U.S. NTP. Mutat. Res. 257, 229-306.

Ashby, J., Callander, R. D., Paton, D., Zeiger, E., and Ratpan, F. (1988). Weak and unexpected mutagenicity to Salmonella of the rat hepatocarcinogen methapyrilene. Environ. Mol. Mutagen 12, 243-252.

Baillie, T. A., Cayen, M. N., Fouda, H., Gerson, R. J., Green, J. D., Grossman, S. J., Klunk, L. J., LeBlanc, B., Perkins, D. G., and Shipley, L. A. (2002). Drug metabolites in safety testing. Toxicol. Appl. Pharmacol. 182, 188-196.

Barlow, S. M., Kozianowski, G., Wurtzen, G., and Schlatter, J. (2001). Threshold of toxicological concern for chemical substances present in the diet. Report of a workshop, 5-6 October 1999, Paris, France. Food Chem. Toxicol. 39, 893-905.

Barlow, S., Renwick, A. G., Kleiner, J., Bridges, J. W., Busk, L., Dybing, E., Edler, L., Eisenbrand, G., Fink-Gremmels, J., Knaap, A., Kroes, R., Liem, D., Muller, D. J., Page, S., Rolland, V., Schlatter, J., Tritscher, A., Tueting, W., and Wurtzen, G. (2006). Risk assessment of substances that are both genotoxic and carcinogenic report of an International Conference organized by EFSA and WHO with support of ILSI Europe. Food Chem. Toxicol. 44, 1636-1650.

Barnes, D. G., and Dourson, M. (1988). Reference dose (RfD): description and use in health risk assessments. Regul. Toxicol. Pharmacol. 8, 471-486.

Beddowes, E. J., Faux, S. P., and Chipman, J. K. (2003). Chloroform, carbon tetrachloride and glutathione depletion induce secondary genotoxicity in liver cells via oxidative stress. Toxicology 187, 101-115.

Benfenati, E., Benigni, R., Demarini, D. M., Helma, C., Kirkland, D., Martin, T. M., Mazzatorta, P., Ouédraogo-Arras, G., Richard, A. M., Schilter, B., Schoonen, W. G. E. J., Snyder, R. D., and Yang, C. (2009). Predictive Models for Carcinogenicity and Mutagenicity: Frameworks, State-of-the-Art, and Perspectives, Vol. 27, pp. 57-90. Taylor & Francis.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. Royal Stat. Soc. B 57, 289-300.

Benz, R. D. (2007). Toxicological and clinical computational analysis and the US FDA/CDER. Expert Opin. Drug Metab. Toxicol. 3, 109-124.

Blackburn, K., Stickney, J. A., Carlson-Lynch, H. L., McGinnis, P. M., Chappell, L., and Felter, S. P. (2005). Application of the threshold of toxicological concern approach to ingredients in personal and household care products. Regul. Toxicol. Pharmacol. 43, 249-259.

Board on Environmental Studies and Toxicology (2008). Science and Decisions: Advancing Risk Assessment. The National Academies Press.

Boobis, A. R., Cohen, S. M., Dellarco, V., McGregor, D., Meek, M. E., Vickers, C., Willcocks, D., and Farland, W. (2006). IPCS framework for analyzing the relevance of a cancer mode of action for humans. Crit. Rev. Toxicol. 36, 781-792.

Brazma A, Parkinson H, Sarkans U, Shojatalab M, Vilo J, Abeygunawardena N, Holloway E, Kapushesky M, Kemmeren P, Lara GG, Oezcimen A, Rocca-Serra P, Sansone SA. (2003). ArrayExpress--a public repository for microarray gene expression data at the EBI. Nucleic Acids Res. 31, 68-71.

Butterworth, B. E. (2006). A classification framework and practical guidance for establishing a mode of action for chemical carcinogens. Regul. Toxicol. Pharmacol. 45, 9-23.

Butterworth, B. E., and Bogdanffy, M. S. (1999). A comprehensive approach for integration of toxicity and cancer risk assessments. Regul. Toxicol. Pharmacol. 29, 23-36.

Butterworth, B. E., Conolly, R. B., and Morgan, K. T. (1995). A strategy for establishing mode of action of chemical carcinogens as a guide for approaches to risk assessments. Cancer Lett. 93, 129-146.

Cheeseman, M. A., Machuga, E. J., and Bailey, A. B. (1999). A tiered approach to threshold of regulation. Food Chem. Toxicol. 37, 387-412.

Clewell, H. (2005). Use of mode of action in risk assessment: past, present, and future. Regul. Toxicol. Pharmacol. 42, 3-14.

Committee for Medicinal Products for Human Use, European Medicines Agency (2006). Guideline on the limits of genotoxic impurities. EMEA/CHMP/QWP/251344/2006.

Committee for Medicinal Products for Human Use, European Medicines Agency (2008a). Question & Answers on the CHMP Guideline on the Limits of Genotoxic Impurities. EMEA/CHMP/SWP/431994/2007.

Committee for Medicinal Products for Human Use, European Medicines Agency (2008b). Questions and answers on the follow-up to the contamination of Viracept (nelfinavir) with ethyl mesilate. EMEA/CHMP/375807/2008.

Conolly, R. B. (1995). Cancer and non-cancer risk assessment: not so different if you consider mechanisms. Toxicology. 102, 179-188.

Crump, K. S. (1984). A new method for determining allowable daily intakes. Fundam. Appl. Toxicol. 4, 854-871.

Cunningham, M. L. (1996). Role of increased DNA replication in the carcinogenic risk of nonmutagenic chemical carcinogens. Mutat. Res. 365, 59-69.

Cunningham, M. L., Pippin, L. L., Anderson, N. L., and Wenk, M. L. (1995). The hepatocarcinogen methapyrilene but not the analog pyrilamine induces sustained hepatocellular replication and protein alterations in F344 rats in a 13-week feed study. Toxicol. Appl. Pharmacol. 131, 216-223.

Decordier, I., Dillen, L., Cundari, E., and Kirsch-Volders, M. (2002). Elimination of micronucleated cells by apoptosis after treatment with inhibitors of microtubules. Mutagenesis 17, 337-344.

Delaney, E. J. (2007). An impact analysis of the application of the threshold of toxicological concern concept to pharmaceuticals. Regul. Toxicol. Pharmacol. 49, 107-124.

Delker, D., Hatch, G., Allen, J., Crissman, B., George, M., Geter, D., Kilburn, S., Moore, T., Nelson, G., Roop, B., Slade, R., Swank, A., Ward, W., and DeAngelo, A. (2006). Molecular biomarkers of oxidative stress associated with bromate carcinogenicity. Toxicology 221, 158-165.

Dennis, G., Jr., Sherman, B. T., Hosack, D. A., Yang, J., Gao, W., Lane, H. C., and Lempicki, R. A. (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol. 4, P3.

Doak, S. H., Jenkins, G. J., Johnson, G. E., Quick, E., Parry, E. M., and Parry, J. M. (2007). Mechanistic influences for mutation induction curves after exposure to DNA-reactive carcinogens. Cancer Res. 67, 3904-3911.

Dobo, K. L., Greene, N., Cyr, M. O., Caron, S., and Ku, W. W. (2006). The application of structure-based assessment to support safety and chemistry diligence to manage genotoxic impurities in active pharmaceutical ingredients during drug development. Regul. Toxicol. Pharmacol. 44, 282-293.

Dolan, D. G., Naumann, B. D., Sargent, E. V., Maier, A., and Dourson, M. (2005). Application of the threshold of toxicological concern concept to pharmaceutical manufacturing operations. Regul. Toxicol. Pharmacol. 43, 1-9.

Dourson, M. (1996). Uncertainty factors in noncancer risk assessment. Regul. Toxicol. Pharmacol. 24, 107.

Dourson, M. L. (1993). Reference dose of the United States Environmental Protection Agency. Scand. J. Work Environ. Health 19 Suppl 1, 115-118.

Dourson, M. L., and Stara, J. F. (1983). Regulatory history and experimental support of uncertainty (safety) factors. Regul. Toxicol. Pharmacol. 3, 224-238.

Dourson, M. L., Felter, S. P., and Robinson, D. (1996). Evolution of science-based uncertainty factors in noncancer risk assessment. Regul. Toxicol. Pharmacol. 24, 108-120.

Dybing, E., Doe, J., Groten, J., Kleiner, J., O'Brien, J., Renwick, A. G., Schlatter, J., Steinberg, P., Tritscher, A., Walker, R., and Younes, M. (2002). Hazard characterisation of chemicals in food and diet. dose response, mechanisms and extrapolation issues. Food Chem. Toxicol. 40, 237-282.

Ellinger-Ziegelbauer, H., Stuart, B., Wahle, B., Bomann, W., and Ahr, H. J. (2005). Comparison of the expression profiles induced by genotoxic and nongenotoxic carcinogens in rat liver. Mutat. Res. 575, 61-84.

Felter, S., Lane, R. W., Latulippe, M. E., Llewellyn, G. C., Olin, S. S., Scimeca, J. A., and Trautman, T. D. (2009). Refining the threshold of toxicological concern (TTC) for risk prioritization of trace chemicals in food. Food Chem. Toxicol. 47, 2236-2245.

Fielden, M. R., Brennan, R., and Gollub, J. (2007). A gene expression biomarker provides early prediction and mechanistic assessment of hepatic tumor induction by nongenotoxic chemicals. Toxicol. Sci. 99, 90-100.

Fiori, J. M., and Meyerhoff, R. D. (2002). Extending the threshold of regulation concept: de minimis limits for carcinogens and mutagens. Regul. Toxicol. Pharmacol. 35, 209-216.

Foster, W. R., Chen, S. J., He, A., Truong, A., Bhaskaran, V., Nelson, D. M., Dambach, D. M., Lehman-McKeeman, L. D., and Car, B. D. (2007). A retrospective analysis of toxicogenomics in the safety assessment of drug candidates. Toxicol. Pathol. 35, 621-635.

Foth, H., and Hayes, A. (2008). Concept of REACH and impact on evaluation of chemicals. Hum. Exp. Toxicol. 27, 5-21.

Fung, V. A., Barrett, J. C., and Huff, J. (1995). The carcinogenesis bioassay in perspective: application in identifying human cancer hazards. Environ. Health Perspect. 103, 680-683.

Galloway, S. M., Miller, J. E., Armstrong, M. J., Bean, C. L., Skopek, T. R., and Nichols, W. W. (1998). DNA synthesis inhibition as an indirect mechanism of chromosome aberrations: comparison of DNA-reactive and non-DNA-reactive clastogens. Mutat. Res. 400, 169-186.

Gaylor, D. W., and Gold, L. S. (1995). Quick estimate of the regulatory virtually safe dose based on the maximum tolerated dose for rodent bioassays. Regul. Toxicol. Pharmacol. 22, 57-63.

Gerber, C., and Toelle, H. G. (2009). What happened: the chemistry side of the incident with EMS contamination in Viracept tablets. Toxicol. Lett. 190, 248-253.

Glatt, H. (2000). Sulfotransferases in the bioactivation of xenobiotics. Chem. Biol. Interact. 129, 141-170.

Glatt, H., and Meinl, W. (2005). Sulfotransferases and acetyltransferases in mutagenicity testing: technical aspects. Methods Enzymol. 400, 230-249.

Golbraikh, A., and Tropsha, A. (2002). Beware of q2! J. Mol. Graph Model 20, 269-276.

Gold, L. S., Gaylor, D. W., and Slone, T. H. (2003). Comparison of cancer risk estimates based on a variety of risk assessment methodologies. Regul. Toxicol. Pharmacol. 37, 45-53.

Gold, L. S., Manley, N. B., Slone, T. H., and Rohrbach, L. (1999). Supplement to the Carcinogenic Potency Database (CPDB): results of animal bioassays published in the general literature in 1993 to 1994 and by the National Toxicology Program in 1995 to 1996. Environ. Health Perspect. 107 Suppl 4, 527-600.

Gold, L. S., Manley, N. B., Slone, T. H., Rohrbach, L., and Garfinkel, G. B. (2005). Supplement to the Carcinogenic Potency Database (CPDB): results of animal bioassays published in the general literature through 1997 and by the National Toxicology Program in 1997-1998. Toxicol. Sci. 85, 747-808.

Gold, L. S., Sawyer, C. B., Magaw, R., Backman, G. M., de, V. M., Levinson, R., Hooper, N. K., Havender, W. R., Bernstein, L., and Peto, R. (1984). A carcinogenic potency database of the standardized results of animal bioassays. Environ. Health Perspect. 58, 9-319.

Gold, L. S., Slone, T. H., Manley, N. B., Garfinkel, G. B., Hudes, E. S., Rohrbach, L., and Ames, B. N. (1991). The Carcinogenic Potency Database: analyses of 4000 chronic animal cancer experiments published in the general literature and by the U.S. National Cancer Institute/National Toxicology Program. Environ. Health Perspect. 96, 11-15.

Gold, L. S., Slone, T. H., Stern, B. R., Manley, N. B., and Ames, B. N. (1992). Rodent carcinogens: setting priorities. Science. 258, 261-265.

Gombar, V. K., and Enslein, K. (1995). Use of predictive toxicology in the design of new chemicals. In Computer-aided molecular design. Applications in agrochemicals, materials, and pharmaceuticals (C. H. Reynolds, M. K. Holloway and H. K. Cox, eds.), pp. 236-249. ACS, Washington, DC.

Gombar, V. K., and Zhang, H. (2007). VISDOM —Visual In Silico Design Of Molecules —a web-based predictive modeling tool. (In- house software, Eli Lilly and Co., Indianapolis, IN 46285).

Gombar, V. K., Mattioni, B. E., Zwickl, C., and Deahl, T. (2007). Computational approaches for assessment of toxicity—A historical perspective and current status. In Computational toxicology: Risk assessment for pharmaceuticals and environmental chemicals (S. Elkins, ed.), pp. 183-215. John Wiley & Sons, Inc., Hoboken, NJ.

Gombar, V. K., Polli, J. W., Humphreys, J. E., Wring, S. A., and Serabjit-Singh, C. S. (2004). Predicting P-glycoprotein substrates by a quantitative structure-activity relationship model. J. Pharm. Sci. 93, 957-968.

Gonzalez, F. J., Peters, J. M., and Cattley, R. C. (1998). Mechanism of action of the nongenotoxic peroxisome proliferators: role of the peroxisome proliferator-activator receptor alpha. J. Natl. Cancer Inst. 90, 1702-1709.

Guengerich, F. P., Dannan, G. A., Wright, S. T., Martin, M. V., and Kaminsky, L. S. (1982). Purification and characterization of liver microsomal cytochromes p-450: electrophoretic, spectral, catalytic, and immunochemical properties and inducibility of eight isozymes isolated from rats treated with phenobarbital or beta-naphthoflavone. Biochemistry 21, 6019-6030.

Hall, L., Mohney, B., and Kier, L. (1991). The electrotopological state: Structure information at the atomic level. J. Chem. Inf. Comput. Sci. 31, 76-82.

Haseman, J. K. (2000). Using the NTP database to assess the value of rodent carcinogenicity studies for determining human cancer risk. Drug Metab. Rev. 32, 169-186.

Hayashi, M., Kamata, E., Hirose, A., Takahashi, M., Morita, T., and Ema, M. (2005). In silico assessment of chemical mutagenesis in comparison with results of Salmonella microsome assay on 909 chemicals. Mutat. Res. 588, 129-135.

Hogg, R. V., McKean, J. W., and Craig, A. T. (2005). Introduction to mathematical statistics. 6th ed. Pearson Education, Inc. NJ.

Huang da, W., Sherman, B. T., and Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 4, 44-57.

Humfrey, C. D. (2007). Recent developments in the risk assessment of potentially genotoxic impurities in pharmaceutical drug substances. Toxicol Sci. 100, 24-28.

Humphreys, W. G., and Unger, S. E. (2006). Safety assessment of drug metabolites: Characterization of chemically stable metabolites. Chem. Res. Toxicol. 19, 1564-1569.

International Conference on Harmonisation (1996). S2A : Guideline for Industry Specific Aspects of Regulatory Genotoxicity Tests for Pharmaceuticals.

International Conference on Harmonisation (1997a). S2B: Genotoxicity: A standard battery for genotoxicity testing for pharmaceuticals.

International Conference on Harmonisation (1997b). S1B: Testing for carcinogenicity of pharmaceuticals.

International Conference on Harmonisation (1997c). Q3C Guideline Residual Solvents. 9.1 Supplement.

International Conference on Harmonisation (1997d). M3(R1): Maintenance of the ICH Guideline on Non-Clinical Safety Studies for the Conduct of Human Clinical Trials for Pharmaceuticals.

International Conference on Harmonisation (2006a). Q3A(R2): Impurities in new drug substances.

International Conference on Harmonisation (2006b). Q3B(R2): Impurities in new drug products

Iype, P. T., Ray-Chaudhuri, R., Lijinsky, W., and Kelley, S. P. (1982). Inability of methapyrilene to induce sister chromatid exchanges in vitro and in vivo. Cancer Res. 42, 4614-4618.

Jacobson-Kram, D., and Jacobs, A. (2005). Use of genotoxicity data to support clinical trials or positive genetox findings on a candidate pharmaceutical or impurity .... now what? Int. J. Toxicol. 24, 129-134.

Jacobson-Kram, D., and McGovern, T. (2007). Toxicological overview of impurities in pharmaceutical products. Adv. Drug Deliv. Rev. 59, 38-42.

Jenkins, G. J., Doak, S. H., Johnson, G. E., Quick, E., Waters, E. M., and Parry, J. M. (2005). Do dose response thresholds exist for genotoxic alkylating agents? Mutagenesis 20, 389-398.

Kaldor, J. M., and L'Abbe, K. A. (1990). Interaction between human carcinogens. IARC Sci. Publ., 35-43.

Kalgutkar, A. S., Gardner, I., Obach, R. S., Shaffer, C. L., Callegari, E., Henne, K. R., Mutlib, A. E., Dalvie, D. K., Lee, J. S., Nakai, Y., O'Donnell, J. P., Boer, J., and Harriman, S. P. (2005). A comprehensive listing of bioactivation pathways of organic functional groups. Curr. Drug Metab. 6, 161-225.

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 28, 27-30.

Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K. F., Itoh, M., Kawashima, S., Katayama, T., Araki, M., and Hirakawa, M. (2006). From genomics to chemical genomics: new developments in KEGG. Nucleic Acids Res. 34, D354-D357.

Kasper, P. (2004). Assessment and Acceptance of Thresholds of Genotoxic Impurities in New Drug Substances A Regulatory Perspectice. Int. J. Pharm. Med. 18, 209-214.

Kim, B. S., and Margolin, B. H. (1999). Prediction of rodent carcinogenicity utilizing a battery of in vitro and in vivo genotoxicity tests. Environ. Mol. Mutagen 34, 297-304.

Kirkland, D., Aardema, M., Henderson, L., and Müller, L. (2005). Evaluation of the ability of a battery of three in vitro genotoxicity tests to discriminate rodent carcinogens and non-carcinogens I. Sensitivity, specificity and relative predictivity. Mutat. Res. 584, 1-256.

Kirkland, D., Aardema, M., Müller, L., and Makoto, H. (2006). Evaluation of the ability of a battery of three in vitro genotoxicity tests to discriminate rodent carcinogens and non-carcinogens II. Further analysis of mammalian cell results, relative predictivity and tumour profiles. Mutat. Res. 608, 29-42.

Klaunig, J. E., Babich, M. A., Baetcke, K. P., Cook, J. C., Corton, J. C., David, R. M., DeLuca, J. G., Lai, D. Y., McKee, R. H., Peters, J. M., Roberts, R. A., and Fenner-Crisp, P. A. (2003). PPARalpha agonist-induced rodent tumors: modes of action and human relevance. Crit. Rev. Toxicol. 33, 655-780.

Klaunig, J. E., Kamendulis, L. M., and Xu, Y. (2000). Epigenetic mechanisms of chemical carcinogenesis. Hum. Exp. Toxicol. 19, 543-555.

Klopman, G. (1984). Artificial intelligence approach to structure-activity studies. Computer automated structure evaluation of biological activity of organic molecules. J. Am. Chem. Soc. 106, 7315-7321.

Klopman, G. (1992). MULTICASE. 1. A hierarchical computer automated structure evaluation program. Quant. Struct-Act. Relat. 11, 176-184.

Kodell, R. L., and Chen, J. J. (1994). Reducing conservatism in risk estimation for mixtures of carcinogens. Risk. Anal. 14, 327-332.

Krewski, D., and Thomas, R. D. (1992). Carcinogenic mixtures. Risk Anal. 12, 105-113.

Krewski, D., Szyszkowicz, M., and Rosenkranz, H. (1990). Quantitative factors in chemical carcinogenesis: variation in carcinogenic potency. Regul. Toxicol. Pharmacol. 12, 13-29.

Kroes, R., and Kozianowski, G. (2002). Threshold of toxicological concern (TTC) in food safety assessment. Toxicol. Lett. 127, 43-46.

Kroes, R., Galli, C., Munro, I., Schilter, B., Tran, L., Walker, R., and Wurtzen, G. (2000). Threshold of toxicological concern for chemical substances present in the diet: a practical tool for assessing the need for toxicity testing. Food Chem. Toxicol. 38, 255-312.

Kroes, R., Kleiner, J., and Renwick, A. (2005). The threshold of toxicological concern concept in risk assessment. Toxicol. Sci. 86, 226-230.

Kroes, R., Renwick, A. G., Cheeseman, M., Kleiner, J., Mangelsdorf, I., Piersma, A., Schilter, B., Schlatter, J., van, S. F., Vos, J. G., and Wurtzen, G. (2004). Structure-based thresholds of toxicological concern (TTC): guidance for application to substances present at low levels in the diet. Food Chem. Toxicol. 42, 65-83.

Ku, W. W., Bigger, A., Brambilla, G., Glatt, H., Gocke, E., Guzzie, P. J., Hakura, A., Honma, M., Martus, H. J., Obach, R. S., and Roberts, S. (2007). Strategy for genotoxicity testing--metabolic considerations. Mutat. Res. 627, 59-77.

Lai, D. Y. (2004). Rodent carcinogenicity of peroxisome proliferators and issues on human relevance. J. Environ. Sci. Health C Environ. Carcinog. Ecotoxicol. Rev. 22, 37-55.

Lampe, M. A., and Kammerer, R. C. (1990). Species differences in the metabolism and macromolecular binding of methapyrilene: a comparison of rat, mouse and hamster. Xenobiotica 20, 1269-80.

Lee, H. K., Kim, Y. H., and Roh, J. K. (1994). Clastogenicity of methapyrilene hydrochloride in cultured Chinese hamster cells. Mutat. Res. 341, 77-82.

Lijinsky, W. (1984). Chronic toxicity tests of pyrilamine maleate and methapyrilene hydrochloride in F344 rats. Food Chem. Toxicol. 22, 27-30.

Lijinsky, W., and Kovatch, R. M. (1986). Carcinogenicity studies of some analogs of the carcinogen methapyrilene in F344 rats. J. Cancer Res. Clin. Oncol. 112, 57-60.

Lijinsky, W., Kovatch, R. M., and Thomas, B. J. (1992). The carcinogenic effect of methapyrilene combined with nitrosodiethylamine given to rats in low doses. Carcinogenesis 13, 1293-1297.

Lijinsky, W., Reuber, M. D., and Blackwell, B. N. (1980). Liver tumors induced in rats by oral administration of the antihistaminic methapyrilene hydrochloride. Science 209, 817-819.

Lipinski, C. A. (2000). Drug-like properties and the causes of poor solubility and poor permeability. J. Pharmacol. Toxicol. Methods 44, 235-249.

Luffer-Atlas, D. (2008). Unique/major human metabolites: why, how, and when to test for safety in animals. Drug Metab. Rev. 40, 447-463.

Lynch, A., Harvey, J., Aylott, M., Nicholas, E., Burman, M., Siddiqui, A., Walker, S., and Rees, R. (2003). Investigations into the concept of a threshold for topoisomerase inhibitor-induced clastogenicity. Mutagenesis 18, 345-353.

Matthews, E. J., and Contrera, J. F. (1998). A new highly specific method for predicting the carcinogenic potential of pharmaceuticals in rodents using enhanced MCASE QSAR-ES software. Regul. Toxicol. Pharmacol. 28, 242-264.

Matthews, E. J., and Contrera, J. F. (2007). In silico approaches to explore toxicity end points: issues and concerns for estimating human health effects. Expert Opin. Drug Metab. Toxicol. 3, 125-134.

Matthews, E. J., Kruhlak, N. L., Benz, R. D., Contrera, J. F., and Marchant, C. A. (2008). Combined use of MC4PC, MDL-QSAR, BioEpisteme, Leadscope PDM, and Derek for Windows software to achieve high performance, high confidence, mode of action-based predictions of chemical carcinogenesis in rodents. Toxicol. Mechan. Methods 18, 189-206.

Matthews, E. J., Kruhlak, N. L., Cimino, M. C., Benz, R. D., and Contrera, J. F. (2006a). An analysis of genetic toxicity, reproductive and developmental toxicity, and carcinogenicity data: I. Identification of carcinogens using surrogate endpoints. Regul. Toxicol. Pharmacol. 44, 83-96.

Matthews, E. J., Kruhlak, N. L., Cimino, M. C., Benz, R. D., and Contrera, J. F. (2006b). An analysis of genetic toxicity, reproductive and developmental toxicity, and carcinogenicity data: II. Identification of genotoxicants, reprotoxicants, and carcinogens using in silico methods. Regul. Toxicol. Pharmacol. 44, 97-110.

Matthews, E. J., Ursem, C. J., Kruhlak, N. L., Benz, R. D., Sabaté, D. A., Yang, C., Klopman, G., and Contrera, J. F. (2009). Identification of structure-activity relationships for adverse effects of pharmaceuticals in humans: Part B. Use of (Q)SAR systems for early detection of drug-induced hepatobiliary and urinary tract toxicities. Regul. Toxicol. Pharmacol. 54, 23-42.

Mayer, J., Cheeseman, M. A., and Twaroski, M. L. (2008). Structure-activity relationship analysis tools: Validation and applicability in predicting carcinogens. Regul. Toxicol. Pharmacol. 50, 50-58.

McGovern, T., and Jacobson-Kram, D. (2006). Regulation of genotoxic and carcinogenic impurities in drug substances and products. Trends Anal. Chem. 25, 790-795.

Melnick, R. L., Kohn, M. C., and Portier, C. J. (1996). Implications for risk assessment of suggested nongenotoxic mechanisms of chemical carcinogenesis. Environ. Health Perspect. 104 Suppl 1, 123-134.

Mercer, A. E., Regan, S. L., Hirst, C. M., Graham, E. E., Antoine, D. J., Benson, C. A., Williams, D. P., Foster, J., Kenna, J. G., and Park, B. K. (2009). Functional and toxicological consequences of metabolic bioactivation of methapyrilene via thiophene S-oxidation: Induction of cell defence, apoptosis and hepatic necrosis. Toxicol. Appl. Pharmacol. 239, 297-305.

Mirsalis, J. C. (1987). Genotoxicity, toxicity, and carcinogenicity of the antihistamine methapyrilene. Mutat. Res. 185, 309-317.

Müller, L., Mauthe, R. J., Riley, C. M., Andino, M. M., Antonis, D. D., Beels, C., DeGeorge, J., De Knaep, A. G., Ellison, D., Fagerland, J. A., Frank, R., Fritschel, B., Galloway, S., Harpur, E., Humfrey, C. D., Jacks, A. S., Jagota, N., Mackinnon, J., Mohan, G., Ness, D. K., O'Donovan, M. R., Smith, M. D., Vudathala, G., and Yotti, L. (2006). A rationale for determining, testing, and controlling specific impurities in pharmaceuticals that possess potential for genotoxicity. Regul. Toxicol. Pharmacol. 44, 198-211.

Mulder, G. J., Hinson, J. A., Nelson, W. L., and Thorgeirsson, S. N. (1977). Role of sulfotransferase from rat liver in the mutagenicity of N-hydroxy-2-acetylaminofluorene in Salmonella typhimurium. Biochem. Pharmacol. 26, 1356-1358.

Muller, L., and Singer, T. (2009). EMS in Viracept--the course of events in 2007 and 2008 from the non-clinical safety point of view. Toxicol. Lett. 190, 243-247.

Munro, I. C., Kennepohl, E., and Kroes, R. (1999). A procedure for the safety evaluation of flavouring substances. Joint FAO/WHO Expert Committee on Food Additives. Food Chem. Toxicol. 37, 207-232.

Nie, A. Y., McMillian, M., Parker, J. B., Leone, A., Bryant, S., Yieh, L., Bittner, A., Nelson, J., Carmen, A., Wan, J., and Lord, P. G. (2006). Predictive toxicogenomics approaches reveal underlying molecular mechanisms of nongenotoxic carcinogenicity. Mol. Carcinog. 45, 914-933.

Obach, R. S., and Dobo, K. L. (2008). Comparison of metabolite profiles generated in Aroclor-induced rat liver and human liver subcellular fractions: considerations for in vitro genotoxicity hazard assessment. Environ. Mol. Mutagen 49, 631-641.

Oberly, T. J., Scheuring, J. C., Richardson, K. A., Richardson, F. C., and Garriott, M. L. (1993). The evaluation of methapyrilene for bacterial mutation with metabolic activation by Aroclor-induced, methapyrilene-induced and noninduced rat-liver S9. Mutat. Res. 299, 77-84.

Package Insert (2006). Triglide (fenofibrate) tablet. First Horizon Pharmaceutical.

Peto, R., Pike, M. C., Bernstein, L., Gold, L. S., and Ames, B. N. (1984). The TD50: a proposed general convention for the numerical description of the carcinogenic potency of chemicals in chronic-exposure animal experiments. Environ. Health Perspect. 58, 1-8.

Pierson, D. A., Olsen, B. A., Robbins, D. K., DeVries, K. M., and Varie, D. L. (2009). Approaches to Assessment, Testing Decisions, and Analytical Determination of Genotoxic Impurities in Drug Substances. J. Org. Proc. Res. Dev. 13, 285-291.

Pozniak, A., Muller, L., Salgo, M., Jones, J. K., Larson, P., and Tweats, D. (2009). Elevated ethyl methanesulfonate (EMS) in nelfinavir mesylate (Viracept, Roche): overview. AIDS Res. Ther. 6, 18.

Renwick, A. G. (2005). Structure-based thresholds of toxicological concern-guidance for application to substances present at low levels in the diet. Toxicol. Appl. Pharmacol. 207, 585-591.

Saeed, A. I., Bhagabati, N. K., Braisted, J. C., Liang, W., Sharov, V., Howe, E. A., Li, J., Thiagarajan, M., White, J. A., and Quackenbush, J. (2006). TM4 microarray software suite. Methods Enzymol. 411, 134-193.

Saeed, A. I., Sharov, V., White, J., Li, J., Liang, W., Bhagabati, N., Braisted, J., Klapa, M., Currier, T., Thiagarajan, M., Sturn, A., Snuffin, M., Rezantsev, A., Popov, D., Ryltsov, A., Kostukovich, E., Borisovsky, I., Liu, Z., Vinsavich, A., Trush, V., and Quackenbush, J. (2003). TM4: a free, open-source system for microarray data management and analysis. Biotechniques 34, 374-378.

Sawyer, C., Peto, R., Bernstein, L., and Pike, M. C. (1984). Calculation of carcinogenic potency from long-term animal carcinogenesis experiments. Biometrics 40, 27-40.

Seidel, S. D., Kan, H. L., Stott, W. T., Schisler, M. R., and Gollapudi, B. B. (2003). Identification of transcriptome profiles for the DNA-damaging agents bleomycin and hydrogen peroxide in L5178Y mouse lymphoma cells. Environ. Mol. Mutagen 42, 19-25.

Seidel, S. D., Stott, W. T., Kan, H. L., Sparrow, B. R., and Gollapudi, B. B. (2006). Gene expression dose-response of liver with a genotoxic and nongenotoxic carcinogen. Int. J. Toxicol. 25, 57-64.

Smith, D. A., and Obach, R. S. (2005). Seeing through the mist: abundance versus percentage. Commentary on metabolites in safety testing. Drug Metab. Dispos. 33, 1409-1417.

Smith, D. A., and Obach, R. S. (2006). Metabolites and safety: What are the concerns, and how should we address them? Chem. Res. Toxicol. 19, 1570-1579.

Snyder, R. D., Pearl, G. S., Mandakas, G., Choy, W. N., Goodsaid, F., and Rosenblum, I. Y. (2004). Assessment of the sensitivity of the computational programs DEREK, TOPKAT, and MCASE in the prediction of the genotoxicity of pharmaceutical molecules. Environ. Mol. Mutagen 43, 143-158.

Steinmetz, K. L., Tyson, C. K., Meierhenry, E. F., Spalding, J. W., and Mirsalis, J. C. (1988). Examination of genotoxicity, toxicity and morphologic alterations in hepatocytes following in vivo or in vitro exposure to methapyrilene. Carcinogenesis 9, 959-963.

Thomas, R. S., Allen, B. C., Nong, A., Yang, L., Bermudez, E., Clewell, H. J., 3rd, and Andersen, M. E. (2007). A method to integrate benchmark dose estimates with genomic data to assess the functional effects of chemical exposure. Toxicol. Sci. 98, 240-248.

Turner, N. T., Woolley, J. L., Jr., Hozier, J. C., Sawyer, J. R., and Clive, D. (1987). Methapyrilene is a genotoxic carcinogen: studies on methapyrilene and pyrilamine in the L5178Y/TK +/- mouse lymphoma assay. Mutat. Res. 189, 285-297.

Uehara, T., Hirode, M., Ono, A., Kiyosawa, N., Omura, K., Shimizu, T., Mizukawa, Y., Miyagishima, T., Nagao, T., and Urushidani, T. (2008). A toxicogenomics approach for early assessment of potential non-genotoxic hepatocarcinogenicity of chemicals in rats. Toxicology 250, 15-26.

United States Environmental Protection Agency (1986). Guidelines for Mutagenicity Risk Assessment. EPA/630/R-98/003.

United States Environmental Protection Agency (1991). Amendments to the Water Quality Standards Regulation; Compliance With CWA Section 303(c)(2)(B); Proposed Rule. Fed. Regist. 56, 223.

United States Environmental Protection Agency (2000a). Risk Characterization Handbook. EPA100B02.

United States Environmental Protection Agency (2000b). Supplementary guidance for conducting health risk assessment of chemical mixtures. EPA/630/R-00/002.

United States Environmental Protection Agency (2000c). Benchmark Dose Technical Guidance Document (Draft). EPA/630/R-00/001.

United States Environmental Protection Agency (2002). A review of the reference dose and reference concentration processes. EPA/630/P-02/002F.

United States Environmental Protection Agency (2005a). Guidelines for carcinogen risk assessment. EPA/630/P-03/001B.

United States Environmental Protection Agency (2005b). Supplemental Guidance for Assessing Susceptibility from Early-Life Exposure to Carcinogens. EPA/630/R-03/003F.

United States Food and Drug Administration : Center for Drug Evaluation and Research (2005). Guidance for Industry Estimating the Maximum Safe Starting Dose in Initial Clinical Trials for Therapeutics in Adult Healthy Volunteers.

United States Food and Drug Administration: Center for Drug Evaluation and Research (2008a). Guidance for industry: Genotoxic and carcinogenic impurities in drug substances and products: Recommended approaches.

United States Food and Drug Administration: Center for Drug Evaluation and Research (2008b). Guidance for Industry Safety Testing of Drug Metabolites.

Venkatapathy, R., Wang, C. Y., Bruce, R. M., and Moudgal, C. (2009). Development of quantitative structure-activity relationship (QSAR) models to predict the carcinogenic potency of chemicals: I. Alternative toxicity measures as an estimator of carcinogenic potency. Toxicol. Appl. Pharmacol. 234, 209-221.

Votano, J. R., Parham, M., Hall, L. H., Kier, L. B., Oloff, S., Tropsha, A., Xie, Q., and Tong, W. (2004). Three new consensus QSAR models for the prediction of Ames genotoxicity. Mutagenesis 19, 365-377.

Walker, V. E., Casciano, D. A., and Tweats, D. J. (2009). The Viracept-EMS case: impact and outlook. Toxicol. Lett. 190, 333-339.

Waring, J. F., Ulrich, R. G., Flint, N., Morfitt, D., Kalkuhl, A., Staedtler, F., Lawton, M., Beekman, J. M., and Suter, L. (2004). Interlaboratory evaluation of rat hepatic gene expression changes induced by methapyrilene. Environ. Health Perspect. 112, 439-448.

White, A. C., Mueller, R. A., Gallavan, R. H., Aaron, S., and Wilson, A. G. (2003). A multiple in silico program approach for the prediction of mutagenicity from chemical structure. Mutat. Res. 539, 77-89.

Williams, E. S., Panko, J., and Paustenbach, D. J. (2009). The European Union's REACH regulation: a review of its history and requirements. Crit. Rev. Toxicol. 39, 553-575.

Yang, L., Allen, B. C., and Thomas, R. S. (2007). BMDExpress: a software tool for the benchmark dose analyses of genomic data. BMC Genomics 8, 387.

Yu, S., Rao, S., and Reddy, J. K. (2003). Peroxisome proliferator-activated receptors, fatty acid oxidation, steatohepatitis and hepatocarcinogenesis. Curr. Mol. Med. 3, 561-572.

Yu, X., Griffith, W. C., Hanspers, K., Dillman, J. F., 3rd, Ong, H., Vredevoogd, M. A., and Faustman, E. M. (2006). A system-based approach to interpret dose- and time-dependent microarray data: quantitative integration of gene ontology analysis for risk assessment. Toxicol. Sci. 92, 560-577.

Zeiger, E., Ashby, J., Bakale, G., Enslein, K., Klopman, G., and Rosenkranz, H. S. (1996). Prediction of Salmonella mutagenicity. Mutagenesis 11, 471-484.

CURRICULUM VITAE

Joel P. Bercu

Education

Indiana University School of Informatics Indianapolis, IN	PhD	Informatics	2010
University of Texas Houston School of Public Heath Houston, TX	MPH	Toxicology	2001
Texas A&M University College Station, TX	BS	Biomedical Sciences	1998

Job History

Lilly Research Laboratories Eli Lilly and Company Indianapolis, IN	Consultant Toxicologist	2001-present
Lyondell Chemical Company Houston, TX	Toxicologist	1999-2001
University of Texas Houston School of Public Health Houston, TX	Graduate Research Assistant	1998-2001
Moffitt Cancer Center	Research Assistant	1997

Professional Memberships

International Life Sciences Institute - Health Environmental and Sciences Institute
Society of Toxicology
Risk Assessment Specialty Section
Occupational Health Specialty Section

Awards/Honors/Citations

Magna Cum Laude (Texas A&M)
Best and Brightest Finalist for Indianapolis
Golden Key Honor Society (Indiana University)
Delta Omega Public Health Honor Society (University of Texas)
Distinguished Student for Biomedical Sciences (Texas A&M)
Eagle Scout

Publications

Bercu JP, Morton SM, Deahl JT, Gombar VK, Callis CM, van Lier RBL. 2010. In silico approaches to predicting cancer potency for risk assessment of genotoxic impurities in drug substances. Regul. Toxicol. Pharmacol. (In Press).

Bercu JP, Jolly RA, Flagella KM, Baker TK, Romero P, Stevens JL. 2010. Toxicogenomics and cancer risk assessment: A framework for key event analysis and dose-response assessment for nongenotoxic carcinogens. Toxicol. Sci. (Submitted).

Baertschi SW, De Antonis D, McKeown A, Bercu JP, Raillard S, Riley CM. 2010. Stress testing as a predictive tool for the assessment of potential genotoxic degradants. In: Pharmaceutical Stress Testing: Predicting Drug Degradation, 2nd Ed. (Editors: S.W. Baertschi SW, Alsante KM, and Reed RA). Informa Life Sciences (In Press).

Raillard SP, Bercu JP, Riley CM, Baertschi SW. Prediction of drug degradation pathways leading to potential genotoxic impurities. Org. Proc. Res. Dev. (In Press).

Callis CM, Bercu JP, DeVries KM, Dow LK, Robbins DK, Varie DL. 2010. Risk assessment of genotoxic impurities in marketed compounds administered over a short-term duration: Applications to oncology products and implications for impurity control limits. Org. Proc. Res. Dev. (Submitted).

Bercu JP, Dobo KL, Gocke E, McGovern T. 2009. Summary of Symposium: Overview of genotoxic impurities in pharmaceutical development. Int. J. Toxicol. 28(6): 468-78.

Bercu JP, Hoffman WP, Lee C, Ness DK. 2008. Auantitative assessment of cumulative carcinogenic risk for multiple genotoxic impurities in a new drug substance. Regul. Toxicol. Pharmacol. 51(3): 270-7.

Bercu JP, Parke NJ, Fiori JM, Meyerhoff RD. 2008. Human health risk assessments for three neuropharmaceutical compounds in surface waters. Regul. Toxicol. Pharmacol. 50(3): 420-7.

Bercu JP, Callis CM. 2009. Comprehensive toxicology risk assessment for genotoxic impurities. Org. Proc. Res. Dev. 13(5): 938.

Dobo KL, Obach RS, Luffer-Atlas D, Bercu JP. 2009. A strategy for risk assessment for human genotoxic metabolites. Chem. Res. Toxicol. 22(2): 348-56.

Bercu JP. 2001. compounds that bind to the gamma-aminobutyric acid benzodiazepine ionophore complex modulate the cellular response to oxidant stress. Thesis - University of Texas Health Science Center at Houston. School of Public Health.

Abstracts

Bercu JP, Morton SM, Deahl JT, Gombar VK, Callis CM, Zwickl CM, van Lier RL. 2009. Risk assessment of genotoxic impurities in a pharmaceutical using in silico tools. Toxicology and Risk Assessment Conference.

Bercu JP. 2008. Risk assessment of genotoxic and carcinogenic impurities: Current issues and case studies. Int. J. Toxicol. 27(6): 405.

Bercu JP, Dobo KL. 2007. Identification and risk assessment of human genotoxic drug metabolites. The Toxicologist. 96(1): 331.

Callis CM, Bercu JP, Moreno MA. 2007. Toxicology and analytical chemistry: An integrated strategy for leachable studies on a polymeric stopper material. Parenteral Drug Association.

Bercu JP, Callis CM, Fiori JM, Meyerhoff RD. 2006. Process to minimize genotoxic impurities in production of a new drug substance. The Toxicologist. 90(1): 179.

Bercu JP, Fiori JM, Meyerhoff RD. 2004. Human health risk assessment for environmental residues of atomoxetine. The Toxicologist. 78(1-S): 152.

Callis CM, Fiori JM, Bercu JP, Meyerhoff RD. 2004. The Threshold of Toxicological Concern (TTC) as an acceptable limit for leachables from polymeric materials into pharmaceutical ingredients. Parenteral Drug Association.

Bercu JP, Smith MA. 2000. Modulation of the benzodiazepine (BZD) receptor/GABA complex affects cell death induced by oxidant stress in Madin Darby canine kidney (MDCK) cells. The Toxicologist. 54(1): 406.

Invited Presentations

Genetic Toxicology Association. 2009. Managing Genotoxic Impurities throughout Development of Pharmaceutical: Control Strategies and Case Studies.

Workshop Sponsored by International Life Sciences Institute - Health Environmental and Sciences Institute. 2009. Short-Term and Intermittent Exposure to Carcinogens - Current Approaches.

American College of Toxicology. 2009. The Threshold of Toxicological Concern - Background and Applications. Session Chair.

Midwest Biopharmacueticals Statistics Workshop. 2009. Linking CMC and Toxicology: The Use of Cumulative Carcinogenic Risk for Multiple Genotoxic Impurities Criteria.

Occupational Toxicology Roundtable. 2009. Chemical Reaction Appraisal to Identify Plausible Genotoxic Impurities with QSAR.

American College of Toxicology. 2008. Risk Assessment of Genotoxic and Carcinogenic Impurities: Current Issues and Case Studies.

Society of Toxicology; National Meeting. 2008. Using Genetic Toxicology Information to Evaluate Safety of Impurities and Metabolites for Pharmaceuticals.

Toxicology and Risk Assessment Conference. 2008. Risk Management of Genotoxic Compounds in the Manufacture of Pharmaceuticals.

Occupational Toxicology Round Table. 2006. Quantitative Structure Analysis and Genotoxic Impurities.

Gulf Coast Society of Toxicology Meeting. 1999. Modulation of Benzodiazepine (BZD) Receptor/ GABAA Complex Affects Cell Death Induced by Oxidant and Stress in Madin Darbin Kindney (MDCK) Cells.