I SPECIFIC AREAS OF PREDICTIVE TOXICOLOGY

1 The human predictive value of combined animal toxicity testing

Current state and emerging approaches

Harry M. Olson and Thomas S. Davies

1.1 INTRODUCTION

Pharmaceutical development in the late twentieth and early twenty-first centuries has been a challenging enterprise. It is an expensive undertaking with a high degree of risk associated mainly with a high failure rate. A new chemical entity (NCE) that successfully completes the entire process of drug discovery and development reaching approval as a new therapeutic may accrue total development costs in excess of one billion dollars (U.S.).[1] Also, for the drugs that are successful, it typically takes 10 to 12 years from the initiation of research efforts to reach final marketing approval.[1] Experience in the past decade with the overall success/failure rate process – which now encompasses many new and emerging tools including early screening assays and *in silico* technologies, and the historical experience of "what works and doesn't work" – has so far not yielded the expected productivity improvements. Enigmatically, recent experience suggests that it is getting more difficult to identify successful lead molecules that lead to safe and effective therapeutics.

A high-level schematic overview of the current drug development process is shown in Figure 1.1. Candidate molecules entering preclinical development from the discovery process proceed through the stages of clinical development (Clinical Phases I, II, and III). During clinical development, safety (first) and efficacy are evaluated in consecutively larger groups of normal volunteers and patients. First-in-human (or FIH) studies number in the 10s of normal subjects or patients, and then the NCE is assessed in patients (100s in Phase 2, and 1,000s in Phase 3) with the disease of therapeutic interest. This overall "classical" approach is oriented mainly to small synthetic molecules, but it is applicable (with modifications) to other drug categories (e.g., biologicals and botanical products) as well.

The drug discovery phase preceding clinical drug development is when evaluation of many potential NCE molecules occurs and the numbers of promising structures is reduced and refined to a select handful of very promising candidates. The goal in discovery is to apply both pharmacology and toxicology screening processes relevant to the intended therapeutic indication and route of administration to identify candidate molecules with the most favorable efficacy

Clinical Phase	Objective	Subjects	No. Studied
Phase I	Safety, PK	Normal volunteers	10's
Phase II	Safety, Efficacy	Patients	100's
Phase III	Safety, Efficacy	Patients	1000's

Figure 1-1: Toxicology in drug development. Reprinted from *Regulatory Toxicology and Pharmacology*, vol. 32/1, Olson et al., "Concordance of the Toxicity of Pharmaceuticals in Humans and in Animals" 12, 2000, with permission from Elsevier.

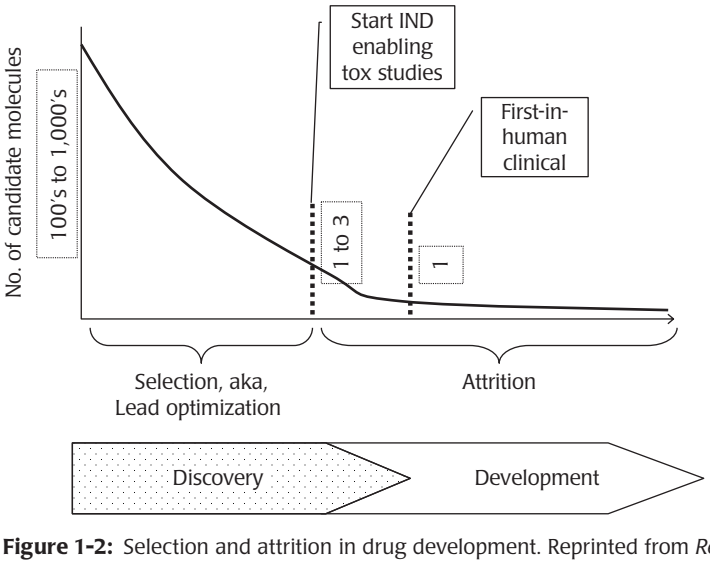

Figure 1-2: Selection and attrition in drug development. Reprinted from *Regulatory Toxicology and Pharmacology*, vol. 32/1, Olson et al., "Concordance of the Toxicity of Pharmaceuticals in Humans and in Animals" 12, 2000, with permission from Elsevier.

and safety attributes so that they can be considered for further development. This process is known as candidate selection (see Figure 1.2), or lead molecule optimization. The technologies and resources applied during candidate selection may include *in silico* methods (e.g., Quantitative Structure Activity Relationship [QSAR]), *in vitro* efficacy and safety screens, and also possibly some *in vivo* animal model assessment (e.g., pharmacokinetic and toxicology screening studies).[2,3]

1.1.1 Candidate selection and attrition – the inevitability of failure

As previously described, candidate selection is the approach in drug discovery that is expected to yield a small and select number of promising molecules for

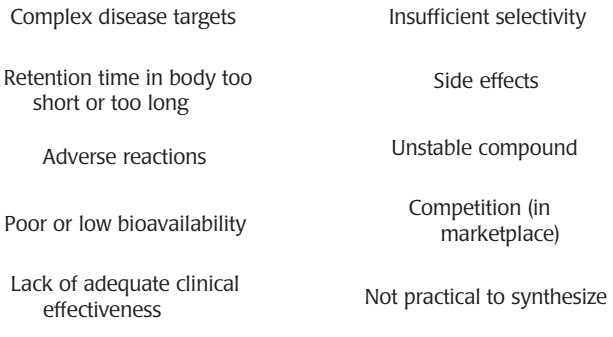

Figure 1-3: Some common causes of attrition. Reprinted from *Regulatory Toxicology and Pharmacology*, vol. 32/1, Olson et al., "Concordance of the Toxicity of Pharmaceuticals in Humans and in Animals" 12, 2000, with permission from Elsevier.

the intended therapeutic indication (Figure 1.2). Many compounds or chemical classes during selection are evaluated in virtual or actual test systems. For toxicologists, this may include *in silico* assessment for structural alerts such as for genotoxic potential, cellular damage, or other potential toxic effects.[2,3] The next step in drug development of the selected lead molecule is attrition. Attrition is the loss of molecules or drug candidates that have entered the preclinical development or subsequent clinical development phases (Figure 1.2). A substantial number of development candidates fail because toxicity issues surfaced in preclinical Investigative New Drug Application (IND)-supporting studies (these compounds selected pre-FIH may never make it into Phase I clinical trials), in subsequent toxicology studies conducted during development, or as a result of significant clinical adverse events arising during development. Attrition can occur even after product registration and marketing, possibly resulting in withdrawal of the product from the market.

Some of the common reasons for attrition are shown in Figure 1.3; these include both toxicity-related findings and other deficiencies in drug candidates that can occur at any time during drug development. Indeed, attrition is about more than toxicity, which nonetheless does remain an important contributor to drug loss.[4] The other factors listed in Figure 1.3 should be kept in mind as key factors in this loss of NCE molecules during development. Unlike during the candidate selection phase, attrition is not a desirable outcome, but it is a normal and expected outcome of drug development. Historically (e.g., in large pharmaceutical company portfolios), attrition can reach or exceed a 90 percent failure rate of compounds identified as "promising" in the late discovery and early development phases. Overall, this may be considered as a kind of Darwinian "natural selection" process, to identify drugability shortcomings early and to focus available resources on the most promising candidates.

Optimally, drug developers want attrition to occur as early as possible, and in particular in the interval between start of the FIH- (or IND)-supporting studies and the end of the Phase II clinical trials (Figure 1.1). Later stage attrition (e.g., in Phase III clinical trials) can have a profound negative impact on drug development programs, timing, and costs.[1] The withdrawal of drugs from the

marketplace due to toxicity or other issues can be discouraging and frustrating (or worse) to patients, drug manufacturers, and regulators.

The understanding of attrition occurring in drug development currently is more grounded historically to small, synthetic molecule development programs than to the more recent advent of biological therapeutics. While there is not yet the long history of experience with biologicals, some manufactured protein/polypeptides are typically "purpose designed" and configured as "humanized" to modulate or replace endogenous molecules, such as insulin, blood clotting factors, or other biomolecules. However, the early stage development of all ethical therapeutics – whether small molecules or biological therapeutics – must include preclinical toxicology evaluation in recognized and accepted (by regulatory authorities) *in vivo* test systems that are recognized as predictive of potential human toxicities or adverse effects.

1.1.2 *In silico*, *in vitro*, and *in vivo* – what approaches to use, and when?

The approach used to identify lead therapeutic molecules must be adaptable to – and take account of – the clinical program therapeutic goals, understand the attributes of the candidate molecule (both unique to the NCE and to the chemical class), and recognize the capabilities and limitations of any test systems used to characterize toxicity. As shown in Figure 1.4, the approach should begin with consideration of *in silico* approaches that may precede *in vitro* and *in vivo* studies.[2,3] Included here are access to the published literature, FOI (freedom-of-information) resources, and archival documentation including computer-based and searchable compound databases, which are important to review what is already known about potential target organ toxic effects for the chemical class. QSAR searchable artificial intelligence systems – a few are Internet accessible – can screen chemical structures and identify possible or "suspect" structural alerts. This information can be useful if applied judiciously to the process of optimizing lead molecule selection. Some *in silico* systems (MultiCASE) are useful to gather information sourced from historical databases and literature references. Many large pharmaceutical companies have systems that include data, summaries, and reports from internal investigative studies.

In vitro screening assays (Figure 1.4) are similarly useful to identify specific molecular characteristics such as receptor affinity, *ex vivo* cross-species comparative metabolite profile, compound metabolic stability, mutagenic potential or other specific attributes or liabilities.[2,3] In preclinical development preceding the Phase I clinical trial, there are specific regulatory requirements for *in vitro* studies conducted under Good Laboratory Practice (GLP) guidances,[3,4] including mutagenic potential in bacteria, structural chromosomal damage in mammalian cells, and also assessment of potassium channel inhibition in the **h**uman ***E**ther-à-go-go* **R**elated **G**ene (hERG) test system.[2–4] Published data from investigative or regulatory *in vitro* systems may provide useful guidance of similar molecule characteristics. *In silico* and *in vitro* evaluations can contribute much to

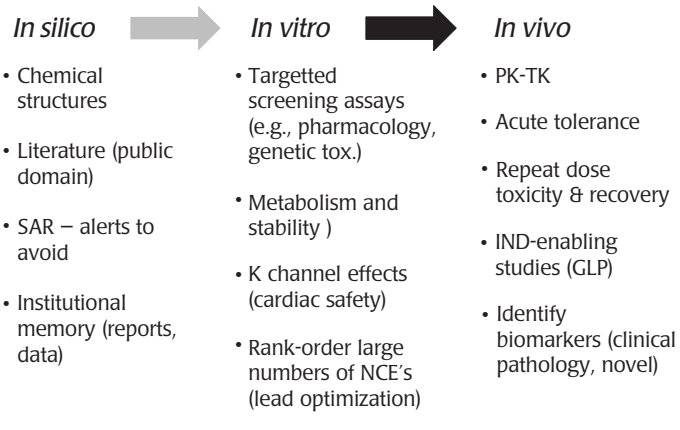

Figure 1-4: Strategic approach to *in vivo* study. Reprinted from *Regulatory Toxicology and Pharmacology*, vol. 32/1, Olson et al., "Concordance of the Toxicity of Pharmaceuticals in Humans and in Animals" 12, 2000, with permission from Elsevier.

the understanding of the molecular toxicity profile, including a preliminary estimate of the probability of success with the NCE development.[2] However, the results from *in vitro* systems have real limitations, and can't be used to reliably characterize the toxicity profile of a selected lead molecule as a therapeutic candidate. For this, the essential resource is *in vivo* toxicity testing.

The inclusion of laboratory animals in studies to assess the toxic potential of lead molecules as a basis for further clinical development is integral to the process of data-based decision making for drug development (Figure 1.4). As an outcome of the cause of death of seventy-six people from diethylene glycol poisoning, as a constituent of a sulphanilamide formulation (Elixir Sulfanilamide), came passage of the federal Food, Drug, and Cosmetic (FD&C) Act of 1938. With this legislation came – for the first time – the requirement of manufacturers of medicines to show that a drug was safe prior to marketing. This and other subsequent legislation mandated the key principles for testing and evaluating new drugs, including the mandate for *in vivo* animal toxicity testing.[5]

Mammalian test systems (notably rat, dog, and primate) are biochemically integrated, with metabolic capabilities for drug transformation and excretion, complex endocrine pathways and feedback loops, and internal organ systems and many clinical pathology biomarkers of toxic effects (e.g., elevated liver enzymes released from damaged hepatocytes or bone marrow toxicity revealed in hemogram changes) that mimic those in humans.[3,6] Therefore, these animal models – most typically one rodent and one nonrodent animal species – have been shown to provide an approximate integrated surrogate to assess possible in-life and target organ toxicity of the NCE. These preclinical studies are important to providing assurance of the safety profile prior to proceeding with studies in human volunteers or patients. Indeed, preclinical *in vivo* models are prime examples of *translation* of toxic effects to human risk

assessment.[6] However, the outcome of toxicology studies performed in these test systems has not always predicted the identical outcomes in humans. But we have acquired much experience about the usefulness of these models to predict human toxicity, and also the limitations to identify certain types of clinical adverse effects.

In order to achieve the end-game of developing a marketable therapeutic, the requisite criteria are that the drug must be shown to be "pure, safe and effective" (Food, Drug and Cosmetic Act of 1938). Practice, experience, and updated regulatory guidance have provided guideposts for meeting each of these three criteria over the course of drug development. The remainder of this chapter will address the extent to which combined animal toxicity testing can help achieve an understanding of potential human toxicity, or adverse effects and what are the shortcomings that remain today.

1.2 MEANING AND VALUE OF PREDICTING HUMAN TOXICITY IN PHARMACEUTICAL DEVELOPMENT

Our attention in this chapter is to focus on contributions to reduce attrition by the pragmatic use of preclinical *in vivo* toxicity testing throughout drug development. This is provided by understanding the use, predictive value, and limitations of preclinical *in vivo* toxicology studies as they pertain to identifying possible human clinical toxicity, and how the safety data obtainable from these preclinical studies is being refined and continues to be improved.

Implicit in the title of this chapter, "The Human Predictive Value of Combined Animal Toxicity Testing," is the expectation *that there is predictive value* in preclinical animal toxicity testing. This is why these studies are done in pharmaceutical research and development, why toxicologists and clinical pharmacologists rely on the data and information from such studies, and, of course, why there are preclinical testing requirements mandated by regulatory authorities. Results from *in vivo* toxicity studies – both the toxicity and toxicokinetic data and their interpretation – are of direct use by clinical investigators to assess risk and benefit of a NCE exposure for patients, particularly in early phases of controlled clinical studies. Support for the predictive value of *in vivo* toxicity testing comes from long experience, historical precedence, published research and regulatory guidance, and requirements spanning many decades. The practice and requirements for animal toxicity design and testing (e.g., using the fewest number of animals needed to obtain a valid scientifically defensible outcome) have been incorporated into current preclinical toxicology study designs. With the availability of current methods to compare *in vitro* metabolism of compounds across species, and to measure plasma levels of parent drug and key (major) metabolites in animal toxicity studies (to compare with therapeutic plasma levels in clinical studies), it is possible to provide expected safety margin estimates for NCEs during development.[2,3]

1.3 *IN VIVO* TESTING STRATEGIES AND MODELS IN-USE FOR DRUG DEVELOPMENT

The background leading to today's preclinical regulatory study testing requirements has been described previously. Guidelines for testing of NCEs are provided in the M-3 Guidance[4] and other preclinical regulatory documents issued by review committees of the International Conferences on Harmonization (ICH). ICH was first convened in 1992 as a cooperative body to reduce duplicate testing of new medicines during the research and development phase. ICH guidance documents and revisions now provide a unified, standardized approach for toxicology testing, achieving greater harmonization of technical guidelines and study designs for product registration.

The main purpose of *in vivo* toxicity testing is to define the preclinical safety profile of the NCE. The safety profile doesn't mean that the drug candidate is absolutely safe in all respects or for all routes of dose administration, but instead that the safety profile of the prospective drug is known. This profile includes in-life effects and target organ toxicity endpoints in relationship to the schedule and route of intended dosing. The systemic drug exposure is reported also in animal toxicology studies. Therefore, "safe" means that the occurrence, incidence, severity, and reversibility of toxic effects and exposure to the NCE in the *in vivo* test system all contribute to determine how the NCE can be administered safely to human subjects. It is important to evaluate and understand the NCE dose-response relationship in studies that include lower doses near to the intended therapeutic exposure in patients, up to higher doses that test the tolerance (toxicology) limits. This dose-response concept was originated by Paracelsus (sixteenth-century "father of toxicology") who advised, "all substances may be poisons, *it is the dose that makes the poison*" (paraphrased, italics the author's).

Therefore, the safety profile of a NCE may include tolerance evaluation, clinical effects, blood or urine biomarkers that signal damage to target organs, and histopathology, which confirms the effect. Toxicokinetic measurements in these studies provide plasma exposure data on the parent compound and/or key metabolites that also occur in humans. During the past decade, the reversibility of toxicity effects is also often evaluated in the study design, by inclusion of a nondosing interval following the dosing phase. Based on this information *a safety margin* can be determined comparing the intended clinical plasma exposure with that in rodent and nonrodent studies to ensure the safety profile is consistent with the intended therapeutic indication. The main objective for drug registration (oncology therapeutics being an exception) is that the drug is expected to be safe under the conditions of intended human investigation and therapeutic use (by route and dose as prescribed).

1.3.1 Predictive value of animal testing

The rationale for laboratory animal testing is that the results from these studies are predictive of possible adverse effects in humans, and therefore can be used to

manage the risk of subsequent human exposure. This is particularly true leading up to initial human trials – also called first-in-human studies – where there is no prior NCE clinical experience.

A very few investigations have been published to determine how predictive animal toxicity testing is related to human toxicities associated with investigative or marketed therapeutics.[7–13] The main aim of these studies was to understand how useful animal toxicity studies are to predict human clinical toxicity. Several of these studies have focused on cytotoxic anticancer therapeutics,[9–12] which have inherently narrow safety margins for tolerable toxicity since higher exposures may engender adverse effects in order to also achieve effective treatment of the disease. For anticancer therapeutics, the observed clinical toxicity in humans may be predicted for at the higher doses in the preclinical toxicity studies because of the very narrow safety margin. For the broader classes of pharmaceuticals (including but not limited to anticancer agents), the predictive value of preclinical animal toxicology studies is addressed by prediction of the clinical toxicity observed during clinical drug development trials (Figure 1.1). Two multinational pharmaceutical company surveys have been undertaken by the International Life Sciences Institute – Human and Environmental Sciences Institute (ILSI-HESI) organization to explore this aspect. The initial survey has been published,[7] and the second survey workshop was held in 2007 [HESI concordance of animal and human toxicity workshop, September 20–21, 2007, Washington, D.C.].

The multinational pharmaceutical company survey is the largest published survey of this kind, including a total of 150 compounds with 221 human toxicities (or adverse effects).[7] Some included drug candidates caused multiple clinical toxicities. In this survey, prediction of the human toxicity was the basis for evaluating whether animals were – or were not – effective to identify the corresponding target organ(s) human toxicity. Schein et al. examined the reported preclinical and human toxicities of twenty-five anticancer drugs in dog, primate, and human studies.[11] Owens reported toxicity findings of twenty-one anticancer drugs in rodent, dog, primate, and human studies,[9] and Freireich and colleagues reported toxicity findings of eighteen anticancer drugs in mouse, rat, hamster, dog, primate, and humans.[12]

The following summarizes the results and conclusions obtained from these published studies for the following organ systems:

Central nervous system (CNS). In a Japanese study of eighty-four drugs evaluated in general pharmacology studies, the reported capacity to predict adverse effects was mixed; however, it reported that changes in locomotor activity in rodents correlated with dizziness in humans.[13] In some studies, high doses in animals produced CNS-related effects (ataxia, convulsions) not observed in clinical trials.[7,14] The concordance in studies of general therapeutics[7] and anticancer therapeutics[9,11] was reported as moderate (predictive from 20 to 60 percent) as there was poor correlation with specific symptoms. Overall, nonrodent data were more predictive than rodent data for identifying adverse neurological effects in the clinic, and histopathologic evaluations were useful to detect serious neurotoxic effects.[15]

Cardiovascular. The overwhelming majority of concordance cases in these studies were reported in nonrodent species, specifically the dog and to a lesser extent the primate.[7,15] For general therapeutics the concordance rate was 80 percent,[7] and this was principally in pharmacology studies. The basis for evaluation includes safety pharmacology electrocardiographic effects and histopathologic toxicities. In these surveys, rodents were determined to have lesser utility because of the unsuitability of this model to evaluate cardiovascular function. Electrocardiographic (ECG) assessment in dogs – combined with *in vitro* assessment in hERG and Purkinje standardized test systems – is currently the regulatory standard to identify compounds presenting higher risk for human cardiac arrhythmias.[2,4]

Hematologic. There was a high concordance (91 percent) in both rodent and nonrodent species with human hematotoxic findings.[7,11,12] These cases are highly correlated with anticancer and antiinfective therapies. Current methods for identifying and evaluating bone marrow toxicity and coagulation effects in both preclinical toxicology and clinical studies are similar, providing the basis for consistent assessment and reliable cross-species and human comparison.

Gastrointestinal (GI). There was a high concordance (85 percent) of human GI toxicities with the animal findings most notably in nonrodent species.[7] This high concordance was particularly the case for anticancer, antiinfective, and antiinflammatory drug classes mediated by pharmacologic mechanisms. Similarly, safety pharmacology studies in a Japanese review of eighty-eight noncancer drugs showed a good correlation for rodent intestinal transport studies and clinical adverse effects (anorexia, constipation).[13] Other studies showed similar positive correlation outcomes across therapeutic classes, including anticancer drugs.[9,11,12,14] In particular the physiologic similarities of the dog and human gastrointestinal tracts may be useful and conducive in support of this high concordance.[16]

Hepatic. Hepatotoxicity remains a significant contributor to attrition in drug development portfolios of many pharmaceutical and biotech companies.[7] Recent surveys indicated that drug-induced liver damage accounts for over 50 percent of incidences of acute liver failure in the United States.[17] In the clinical and preclinical settings, measurement of liver enzymes in blood during NCE dose administration is the most reliable in-life method of detecting potential hepatotoxicity; liver histopathology is also important to confirm severity of effects in animal studies. One study reported a concordance of 80 percent in identifying hepatotoxicity from toxicity studies with known human hepatotoxicity.[18] Other studies assessing anticancer drugs reported good predictivity of hepatotoxic injury in humans with drugs, as evaluated by enzyme changes and liver histopathology.[9,11] In the large multinational industry survey, concordance was shown to be about 50 percent,[7] which was among the lower predictive markers. Possibly this is related to the occurrence of either: (a) subtle preclinical changes (i.e., minimal liver histopathologic changes or low-level increases in liver enzymes in only a few study animals) that fail to be recognized as hepatotoxic signals, or (b) occurrence of idiosyncratic hepatotoxicity with < 1:10,000

incidence in the population. In fact, idiosyncratic reactions are not uncommon and continue to have an impact on late-stage attrition of drugs in development or withdrawals from the marketplace.[15] For large pharmaceutical companies, the occurrence of liver enzyme changes in early drug development in rats is recognized as a common occurrence, generally related to propensity of this species to respond to metabolic inducers. NCEs that show evidence for liver effects and damage are usually presumed to be risky and are dropped in the early screening stages of drug development.

Renal. Renal toxicity – similar to hepatotoxicity – is assessed in preclinical toxicology studies by blood parameters (blood urea nitrogen [BUN], creatinine, electrolytes), urinalysis constituents, and histopathologic examination. Predictive results in studies with anticancer drugs were variable, with a tendency toward overprediction.[9,11] Concordance in the large multinational industry survey was about 70 percent.[7]

Pulmonary. The assessment of drug candidates on respiration is performed prior to FIH registration in either specific safety pharmacology respiratory studies – typically in rodents – or clinical evaluation of rodents and nonrodents in the postdose phase of conventional toxicology studies, and by histopathologic evaluation of lungs. Igarashi et al. reported that respiratory disturbances that occur clinically were not predicted by the safety pharmacology studies.[13] Both Schein and Fletcher reported a high degree of overprediction of respiratory effects in animal toxicity and safety pharmacology studies, compared to actual clinical adverse events.[11,14]

Endocrine. Endocrine changes may be identified by inclusion of specific hormone analyses (based on availability of bioanalytical methods for the species) in toxicology studies, or more typically by histopathologic evaluation of endocrine organs. Results from the multinational industry survey showed only moderate concordance (60 percent) from preclinical studies.[7] Fletcher reported that the findings from preclinical toxicology studies overpredicted effects in humans.[14]

Dermal. There are very few reported cases of skin reaction responses in the multinational industry survey,[7] and animal models in other surveys do not provide reliable predictive utility for these effects.[9,11,14] However, when they occur, dermal hypersensitivity-type reactions are a significant contributor to termination of drugs in various stages of development (Clinical Phases 2 and 3 in particular).[7,19]

Immunologic. The literature is replete with examples of xenobiotic (including therapeutics) immune effects in animal species, but except for hypersensitization few of these effects have been seen in human studies. In many cases immunologic endpoints in animal studies have not been evaluated for predictivity in humans. However, some specific individual species effects that are usefully compared to human immunologic effects (e.g., immune complex effects in rabbits, or immediate/delayed-type hypersensitivity in guinea pigs) have been reported.[20]

A workshop on Preclinical Evaluation of Peptides and Recombinant Proteins provided an integrated interpretation of preclinical toxicology data for