Second Edition

A Comprehensive Guide to **Toxicology in Nonclinical Drug Development**

Edited by
Ali Said Faqi

A COMPREHENSIVE GUIDE TO TOXICOLOGY IN NONCLINICAL DRUG DEVELOPMENT

Dedication

The book is dedicated to my Mom and Dad, to my siblings, my kids (Hani, Abdullahi, Suad and Issra) and to my dear Yasmine Allas

A COMPREHENSIVE GUIDE TO TOXICOLOGY IN NONCLINICAL DRUG DEVELOPMENT

SECOND EDITION

Edited by

ALI SAID FAQI, DVM, PHD, DABT
Fellow ATS Senior Director MPI Research, Mattawan, Michigan, United States
Wayne State University, Detroit, Michigan, United States

ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO
Academic Press is an imprint of Elsevier

Academic Press is an imprint of Elsevier
125 London Wall, London EC2Y 5AS, United Kingdom
525 B Street, Suite 1800, San Diego, CA 92101-4495, United States
50 Hampshire Street, 5th Floor, Cambridge, MA 02139, United States
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, United Kingdom

First edition published as A Comprehensive Guide to Toxicology in Preclinical Drug Development

Notices
Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

ISBN: 978-0-12-803620-4

Publisher: Mica Haley
Acquisition Editor: Kristine Jones
Editorial Project Manager: Molly McLaughlin
Production Project Manager: Laura Jackson
Designer: Victoria Pearson

Typeset by TNQ Books and Journals

Contents

List of Contributors

M.M. Abdi GlaxoSmithKline, Ware, Hertfordshire, United Kingdom

N.H. Al-Humadi Food and Drug Administration (FDA), Silver Spring, MD, United States

C.J. Amuzie[1] Janssen Pharmaceutical Research and Development, Spring House, PA, United States

S.M. Arnold The Dow Chemical Company, Midland, MI, United States

M. Attar Allergan, plc., Irvine, CA, United States

A.D. Aulbach MPI Research, Mattawan, MI, United States

T.J. Baird MPI Research, Mattawan, MI, United States

J.A. Brassard Tustin, CA, United States

T.T.A. Chang Van Andel Research Institute, Grand Rapids, MI, United States

D.B. Colagiovanni Verde Research, Lakewood, CO, United States

J.B. Colerangle Sanofi U.S. Inc., Bridgewater, NJ, United States

R. Collins Biostat Consultants, Portage, MI, United States

J.A. Dalton MPI Research, Mattawan, MI, United States

K.H. Denny Teva Pharmaceuticals, West Chester, PA, United States

R.R. Dietert Cornell University College of Veterinary Medicine, Ithaca, NY, United States

J.N. Duncan Pfizer Inc., Groton, CT, United States

A.S. Faqi MPI Research, Mattawan, MI, United States; Wayne State University, Detroit, MI, United States

J.T. Farmer ICON Development Solutions, Whitesboro, NY, United States

M. Ferrell Ramos AbbVie Inc., North Chicago, IL, United States

P.D. Forbes Toxarus, Inc., Malvern, PA, United States

S. Frantz MPI Research, Mattawan, MI, United States

L. Freshwater Biostat Consultants, Portage, MI, United States

T.C. Fuchs Merck KGaA, Darmstadt, Germany

D.V. Gauvin MPI Research, Mattawan, MI, United States

N.S. Goud Cook Medical, Bloomington, IN, United States

M.D. Green Food and Drug Administration (FDA), Silver Spring, MD, United States

[1]The author transitioned from MPI Research, Mattawan, MI, United States during the publication process.

L. Guo Metabolon Inc., Durham, NC, United States

S.P. Henry Ionis Pharmaceuticals Inc., Carlsbad, CA, United States

P.G. Hewitt Merck KGaA, Darmstadt, Germany

A. Hoberman Charles River Laboratories, Horsham, PA, United States

J.Y. Hui Celgene Corporation, Summit, NJ, United States

C. Johnson Toxicology Consultant, Hamilton, VA, United States

J.W. Kille J.W. Kille Associates, Stanton, NJ, United States

A.S. Kim Theravance Biopharma, South San Francisco, CA, United States

T.W. Kim Ionis Pharmaceuticals Inc., Carlsbad, CA, United States

E. Koch NonClinical Safety Consulting LLC, Denver, CO, United States

D. Kornbrust Preclinsight, Reno, NV, United States

D.B. Learn Charles River Laboratories, Horsham, PA, United States

E. Lewis Charles River Laboratories, Horsham, PA, United States

S. Matsumoto Allergan, plc., Irvine, CA, United States

D.L. McCormick IIT Research Institute, Chicago, IL, United States

O.R. Mendes Eurofins US, Dayton, OH, United States

K.B. Meyer-Tamaki Sangamo BioSciences Inc., Richmond, CA, United States

I. Mikaelian Hoffmann-La Roche, Inc., Nutley, NJ, United States

M.V. Milburn Metabolon Inc., Durham, NC, United States

J. Nicolette AbbVie, Inc., North Chicago, IL, United States

P. Nugent Pfizer Inc., Groton, CT, United States

H.H. Oh Boehringer Ingelheim Pharmaceuticals, Inc., Ridgefield, CT, United States

L. Oyejide Regeneron Pharmaceuticals, Inc., Tarrytown, NY, United States

S. Plassmann PreClinical Safety (PCS) Consultants Ltd, Basel, Switzerland

K. Regal ProPharma Services, LLC, Westminster, CO, United States

D. Rehagen MPI Research, Mattawan, MI, United States

J.C. Resendez Charles River Laboratories, Ashland, OH, United States

J.C. Rowlands The Dow Chemical Company, Midland, MI, United States

J.A. Ryals Metabolon Inc., Durham, NC, United States

C.P. Sambuco Charles River Laboratories (Retired), Horsham, PA, United States

M. Schrag ProPharma Services, LLC, Westminster, CO, United States

R.S. Settivari The Dow Chemical Company, Midland, MI, United States

R. Slauter MPI Research, Mattawan, MI, United States

P.J. Spencer The Dow Chemical Company, Midland, MI, United States

M.E. Stern ImmunEyeZ LLC, Mission Viejo, CA, United States

C.W. Stewart MPI Research, Mattawan, MI, United States

D. Stump WIL Research Laboratories, LLC, Ashland, OH, United States

S. Surapaneni Celgene Corporation, Summit, NJ, United States

M. Templin Preclinical, SNBL USA, Everett, WA, United States

G.L. Truisi Merck KGaA, Darmstadt, Germany

C. Vangyi Glaukos, Laguna Hills, CA, United States

T. Vidmar Biostat Consultants, Portage, MI, United States

J. Vrbanac MPI Research, Mattawan, MI, United States

Z.J. Wang Advance Medical Systems Limited and MITRO Biotech., China

L.O. Whiteley Pfizer Inc., Cambridge, MA, United States

D.M. Wilson The Dow Chemical Company, Midland, MI, United States

J.S. Yan Hutchison MediPharma Ltd., Shanghai, China

M.J. York Bedfordshire, United Kingdom

H.S. Younis NGM Biopharmaceuticals, South San Francisco, CA, United States

Foreword

When the initial edition of *A Comprehensive Guide to Toxicology in Preclinical Drug Development* was published in 2013, I identified the relevance of this new text for those involved in the many different areas of the drug and device development process designed to bring new products into the marketplace, and I indicated that it should become almost mandatory as a resource for those professionals involved in those areas of development. Based on the acceptance of that text by those professionals and with the speed and pace of the science involved in drug and device development, it is not surprising that a second edition is needed so soon after the first. The world we live in seems to change by the minute and we can find ourselves suddenly involved with dangers that we had not contemplated, such as the spread of the deadly Ebola virus in Africa and the nearly worldwide spread of the Zika virus, which has shown to cause severe fetal brain defects such as microcephaly in newborn infants. In response to the Ebola outbreak, we saw a sudden surge by pharmaceutical companies to initiate drug development programs to create vaccines and treatment therapies against this deadly disease. Although the threat of the Zika virus is more recent, we should expect to see the same amount of resources and energy being spent by those in the drug development industry in response to this threat. In each of these cases and in those unknown cases that will occur at some time in the future, the information and guidance of the scientific experts contributing their experience and expertise to the second edition will allow those in the development of new drugs and devices to chart a more predictable and less risky course.

In my opinion, the initial edition of *A Comprehensive Guide to Toxicology in Preclinical Drug Development* quickly became a signal scientific text within the drug development community and remains one of the best, if not the best, comprehensive text currently published dealing with preclinical testing. Its strength lies in its broad yet detailed inclusion of the numerous issues involved with bringing a new drug or device to the marketplace, which includes chapters on toxicogenomics, predictive toxicology, and imaging as new tools to understanding toxicity, and its comprehensiveness in discussing issues associated with both small- and large-molecule compounds. As compared to the 36 chapters in the initial edition, this edition has 35 chapters. Several chapters have either been combined or were identified as being covered in other existing chapters, and most chapters have been revised and updated to reflect the most current state of that particular area of discussion, which is specifically seen in the chapter on juvenile toxicity testing. Four new chapters have been added and I believe that each of these new chapters add relevant and significant completeness to the text. These chapters deal with concepts, strategies, and pitfalls of nonclinical drug development; biomarkers; nonclinical safety assessment of cell-based therapies; and medical devices. Furthermore, the organization of the text has been reordered with a more orderly presentation of the different topics, leading to a better flow among the chapters than in the original edition.

I remain honored that my colleague, Dr. Ali Faqi, has again asked me to write the foreword to this impressive scientific text. Having worked together for over 13 years, I remain fully impressed with his scientific knowledge and personal and professional qualities, and I consider it a rare gift that I can call him my friend.

David G. Serota, Ph.D., D.A.B.T.
Senior Vice President and Director of Research
MPI Research
President, American College of Toxicology - 2012

This page intentionally left blank

CHAPTER

1

Introduction

A.S. Faqi

Drug development is defined as the entire process of bringing a new drug or device to the market. It involves discovery and synthesis, nonclinical development (chemical testing, biological testing, pharmacology, toxicology, safety, etc.), clinical development (Phase I–III), regulatory review, marketing approval, market launch, and postmarketing development (Fig. 1.1).

The process of drug discovery comprises research on (1) target identification, (2) target prioritization/validation, (3) lead identification, and (4) lead optimization.

A range of techniques is used to identify and isolate individual drug targets. The target identification process isolates drugs that have various interactions with the disease targets and might be beneficial in the treatment of a specific disease. A poor understanding of the molecular mechanisms underlying both disease progression and drug action is one of the major challenges of drug discovery as insufficient drug specificity and side effects are often discovered during the late stages of drug development or even after marketing approval [1]. A "target" can be either proteins physically binding to the drug or to proteins that are only functionally related. A decent target needs to be efficacious, safe, meet clinical and commercial needs, and, above all, be "druggable" [2]. On binding, a druggable target elicits a biological response that may be measured both in vitro and in vivo as the putative drug molecule, be that of a small molecule, or biologicals is accessible to the target [2].

Identification of drug target is followed by a target prioritization/validation phase, during which experimental tests are conducted to confirm that interactions with the drug target are associated with the desired change in the behavior of diseased cells. It is the process by which the predicted molecular target of a drug is verified. The following criteria serve as decision-making tools prior to advancement beyond target validation [3]:

1. Known molecules modulating target
2. The target has a history of success
3. Genetic confirmation
4. Availability of known animal models
5. Availability of low-throughput target validation assay that represents biology
6. Intellectual property of the target
7. Determination of the marketability of the target

The next phase following target validation is obviously the lead identification. Identification of lead compounds are sometimes developed as collections, or libraries, of individual molecules that possess the properties required in a new drug. Once the lead is identified, experimental testing is then performed on each of the molecules to confirm their effect on the drug target. This progresses further to lead optimization. Lead optimization studies are conducted on *animals* or in vitro or modulation of a desired target in disease patients [2] to compare various lead compounds, to determine how they are metabolized, and what affect they might induce in the body. The information obtained from lead optimization studies helps scientists in the pharmaceutical industry sort out the compounds with the greatest potential to be developed into a safe and effective drug [2].

Toxicology studies in the drug discovery process are conducted to evaluate the safety of potential drug candidates. This is accomplished using relevant animal models and validated procedures. The ultimate goal is to translate the animal responses into an understanding of the risk for human subjects. This demands additional studies and investment earlier in the candidate evaluation, coupled with an arduous selection process for drug candidates and a speedy kill to avoid spending money and time on compounds that would likely fail in development.

Even after a successful drug candidate for a disease target is identified, drug development still faces enormous challenges, which many drugs fail because of their unacceptable toxicity. The new paradigm in drug discovery should consider a vigorous means of identifying issues related to toxicity early in the discovery process

http://dx.doi.org/10.1016/B978-0-12-803620-4.00001-3

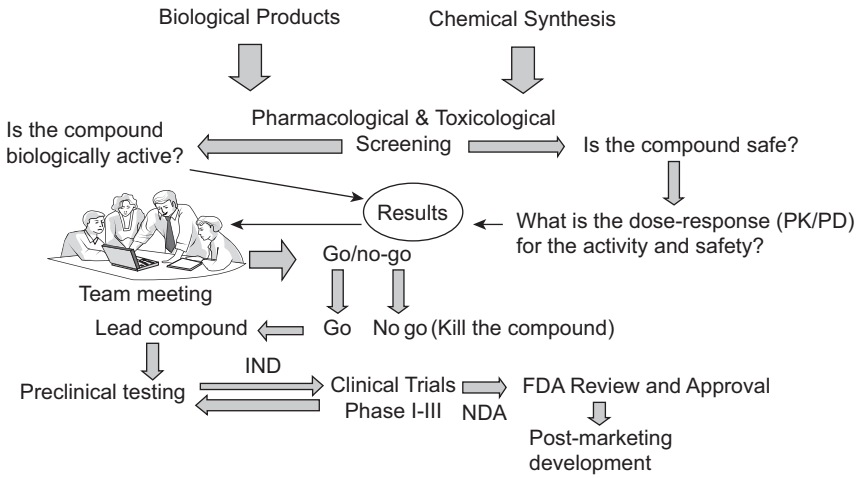

FIGURE 1.1 Drug development and nonclinical testing process.

where the cost of a failed drug is far less than in late drug development stages [4].

The successful drug candidate undergoes a nonclinical safety-testing program. Key factors affecting the type of nonclinical testing include the chemical structure, nature of the compound (small molecules or biologics), proposed human indication, target population, method of administration, and duration of administration (acute, chronic). During nonclinical drug testing, the toxicity and pharmacologic effects of the new chemical entity (NCE) are evaluated by in vitro and in vivo laboratory animal testing. An investigations on drug absorption and metabolism, toxicity of the drug's metabolites, and the speed with which the drug and its metabolites are excreted from the body. Investigational new drug application (IND)-enabling safety studies include acute toxicity of the drug in at least two species of animals, and short-term toxicity studies ranging from 2 weeks to 1 month, genotoxicity, safety pharmacology, and bioanalytical. Furthermore, post-IND nonclinical testing may include subchronic, chronic toxicity, developmental, and reproductive toxicology and carcinogenicity.

It is estimated that it takes 10–12 years to develop and test a new drug before it can be approved for clinical use. This estimate includes early laboratory and animal testing as well as later clinical trials using human subjects. A new study by the Tufts Center for the Study of Drug Development estimates the cost of developing a new drug that gains marketing approval to be around $2.6 billion [5]. Safety issues are the leading cause of attrition at all stages of the drug development process. It is important, however, to understand that the majority of safety-related attrition occurs preclinically, suggesting that approaches that could identify "predictable" nonclinical safety liabilities earlier in the drug development process could lead to the design and/or selection of better drug candidates with increased chances of succeeding for marketing approval [6]. An overview of drug discovery screening assay is shown in Fig. 1.2 [2].

Toxicology testing in animals traditionally focus on phenotypic changes in an organism that result from exposure to the drug; therefore efficient and accurate approaches to assess toxicological effects of drugs on living systems are still less developed. One of the key factors used for a go/no-go decision-making for an NCE relies on the early knowledge of any potential toxic effect. Thus the traditional approach based on the determination of the No-observed-adverse-effect-level (NOAEL) is far from accurate. One of the limitations of this approach is that it may fail to detect adverse effects that manifest at low frequencies.

Indeed, in the past quarter of a century new technologies have emerged that have improved current approaches and are leading to novel predictive approaches for studying disease risk. Increased understanding of the mode of action and the use of scientific tools to predict toxicity is expected to reduce the attrition rate of NCE and thus decrease the cost of developing new drugs. In fact, most big pharmaceutical companies are now using improved model systems for predicting potential drug toxicity, both to decrease the rate of drug-related adverse reactions and to reduce attrition rates. A wide range of biological assay platforms, including toxicogenomics and metabolomics employed in constructing predictive toxicity, are included as separate chapters in this book. The discipline of toxicogenomics is defined as the application of global mRNA, protein and metabolite analysis-related technologies to study the effects of

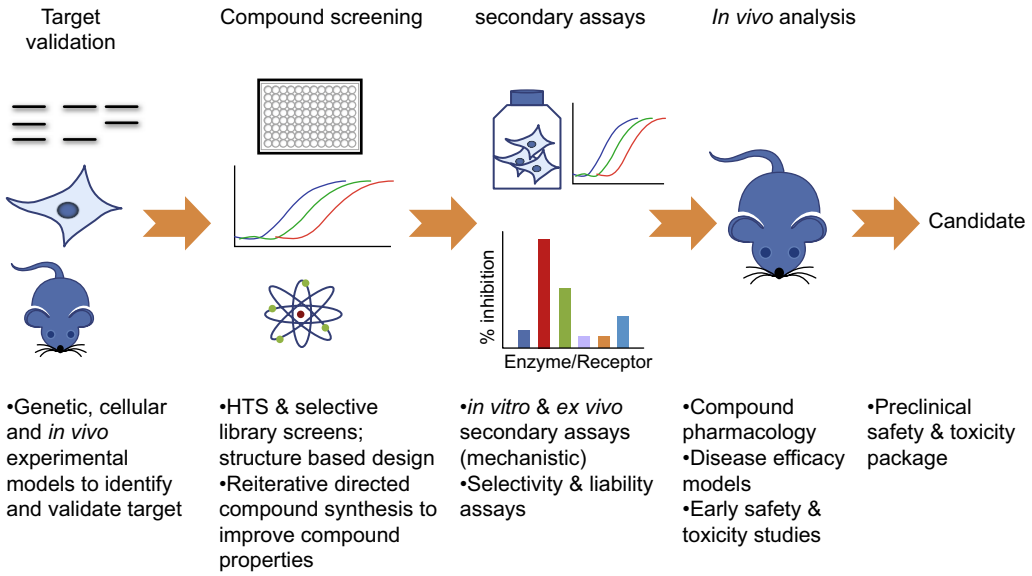

FIGURE 1.2 Overview of drug discovery screening assay. *Hughes JP, Rees S, Kalindjian SB, Philpott KL. Principles of early drug discovery. Br J Pharmacol 2011;162(6):1239–49.*

hazards on organisms [7]. Examining the patterns of altered molecular expression caused by specific exposures can reveal how toxicants act and cause their effect. Identification of toxicity pathways and development of targeted assays to systematically assess potential mode of actions allow for a more thorough understanding of safety issues. Indeed, there is high expectation that toxicogenomics in drug development will predict/better assess potential drug toxicity, and hence reduce failure rates.

In addition metabolomics, a more recent discipline related to proteomics and genomics, uses metabolic signatures to determine the molecular mechanisms of drug actions and predict physiological toxicity. The technology involves rapid and high throughput characterization of the small molecule metabolites found in an organism, and is increasingly gaining attention in nonclinical safety testing. Moreover, the introduction of pharmacogenetics assays has also brought success in drug development in terms of predictability of safety and efficacy. There is a need felt for pharmacogenomics studies, where the effects of multiple genes are assessed with the study of entire genome [8].

Nonclinical safety data are used to select doses in Phase I clinical trial, to provide information on potential side effects, and thus minimize the risk of serious side effects in clinical trials. It also identifies potential target organs and determines toxicity endpoints not amenable to evaluation in clinical trials, such as genetic toxicity, developmental toxicity, and carcinogenicity.

This second edition of the book *Comprehensive Guide to Toxicology in Nonclinical Drug Development* has been reorganized and chapters updated and expanded to include pitfalls in drug development, medical devices, safety assessment of stem cells, and more. Each chapter is carefully crafted to reflect current knowledge and latest research reports/breakthroughs in the field of drug development. The book encompasses series of chapters regrouped into eight units, namely (1) drug discovery, metabolism, and pharmacokinetics, (2) toxicological studies and IND application and first in-human clinical trial, (3) clinical pathology, histopathology, and biomarkers, (4) biostatistics, regulatory toxicology and role of study directors, (5) specialty route of administration, (6) nonclinical development of monoclonal antibodies, stem cells, oncogenic, and nononcogenic drugs, oligonucleotides, and vaccines, (7) safety evaluation of ocular drugs, botanical products, and medicinal devices, and (8) predictive toxicology, toxicometabolomics, toxicogenomics, and imaging.

The book is considered to be a comprehensive guide for toxicologists, regulatory scientists, students, and academics interested in the drug development process and safety testing. It provides a wealth of knowledge of the complex and highly interlinked disciplines of drug development especially in areas of nonclinical safety assessments, absorption, distribution, metabolism, and excretion, regulatory guidelines and submissions, potential pitfalls, biomarkers, predictive toxicology tools, and imaging, which are keys to planning and executing successful drug development projects.

Last but not least, we want to emphasize that one of the biggest strengths of this book comes from the contributors,

who are considered to be leading experts in their respective field. In essence, scientific knowledge gained through experience in the field truly shapes personal lives, hence reinforcing the individual intellect, and wisdom. My expectation is that the second edition will be equally successful as the previous edition if not more.

In conclusion, I would like to thank the contributors for their commitment, hard work, and on time delivery of a high-quality product. Likewise, my deepest gratitude and appreciations goes to Kristine Jones, Molly M. McLaughlin, and Laura Jackson and all the production team at Elsevier.

References

[1] Sivachenko AY, Yuryev A. Pathway analysis software as a tool for drug target selection, prioritization and validation of drug mechanism. Expert Opin Ther Targets 2007;11(3):411–21.

[2] Hughes JP, Rees S, Kalindjian SB, Philpott KL. Principles of early drug discovery. Br J Pharmacol 2011;162(6):1239–49.

[3] Hughes M, Inglese J, Kurtz A, Andalibi A, Patton L, Austin C, et al. Early drug discovery and development guidelines: for academic researchers, collaborators, and start-up companies. In: Sittampalam GS, Coussens NP, Nelson H, et al., editors. A assay guidance manual. 2004. [Bethesda (MD)].

[4] McKim JM. Building a tiered approach to in vitro predictive toxicity screening: a focus on assays with in vivo relevance. Comb Chem High Throughput Screen 2010;13(2):188–206.

[5] http://csdd.tufts.edu/news/complete_story/pr_tufts_csdd_2014_cost_study [accessed 26.02.16].

[6] Kramer JA, Sagartz JE, Morris DL. The application of discovery toxicology and pathology towards the design of safer pharmaceutical lead candidates. Nat Rev Drug Discov 2007;6(8):636–49.

[7] Hamadeh HK, Amin RP, Paules RS, Afshari CA. An overview of toxicogenomics. Curr Issues Mol Biol 2002;4:45–56.

[8] Surendiran A, Pradhan SC, Adithan C. Role of pharmacogenomics in drug discovery and development. Indian J Pharmacol 2008;40(4):137–43.

SECTION I

DRUG DISCOVERY, METABOLISM, AND PHARMACOKINETICS

This page intentionally left blank

CHAPTER

2

Critical Aspects of Integrated Nonclinical Drug Development: Concepts, Strategies, and Potential Pitfalls

E. Koch, S. Plassmann

OUTLINE

A Comprehensive Guide to Toxicology in Nonclinical Drug Development, Second Edition
http://dx.doi.org/10.1016/B978-0-12-803620-4.00002-5

INTRODUCTION

The development of a new drug is a complicated, long, and expensive process and confronts the experts in all disciplines involved with unexpected challenges during the complex process it takes to bring new medicines in a variety of indications to the market. These challenges require integration into what proves to be a demanding journey through often unknown territories that need to be explored as the route unfolds itself on the way.

From understanding a disease and identifying a "druggable" target to bringing a safe and effective new treatment to patients take on average 10–15 years. According to DiMasi et al. [39] the cost to develop and get marketing approval for a new drug is $2.6 billion. This figure includes the price of failure and opportunity costs. The costs have more than tripled from the $800 million estimated for the year 2000 [38]. Of every 5000 to 10,000 compounds entering the research and development pipeline only one may receive approval.

The goal of nonclinical (or preclinical) development is to build a bridge from "bench to bedside," ie, to characterize molecules—small or large—in a step-wise process to support specific phases of clinical development, ie, to generate knowledge providing the basis for the conduct of defined clinical studies of a given scope and duration as well as the marketing authorization for pharmaceuticals. The master guideline describing this process is ICH M3(R2) [90].

However, potential obstacles hindering drug development at any stage are numerous and during each step the unexpected has to be expected. Critical questions need to be addressed along the way, such as what do we understand at a specific moment in time when we are facing an issue? Will this issue be a "show stopper" or only make us take a more complicated but still potentially successful route? How can we mitigate potential risks in humans? How reliably can we translate data from animals or clinical studies to the patient target population? How can we responsibly balance potential risks against essential benefits? How certain can we be as to whether we have asked the right questions along the way? Are we sure we really understand the answers to those questions?

In this chapter, we present historical examples from medicines during different phases of drug development highlighting general or specific issues, that, in turn, may have led either to modifications of general requirements for the nonclinical characterization of investigational medicinal products [IMPs—this term is used in the European Union (EU)]/investigational new drugs (INDs—this term is used in the United States) or to modifications of the requirements for risk management and mitigation in humans or to a combination of both. Certain issues are recognized by the scientific community as unresolved at present, and we must understand ourselves to be inquirers that often will not understand the full picture at a given moment in time as is the intrinsic nature of medicine as an empiric discipline. We will also provide context as to how predictive nonclinical studies for human risk assessment are in certain areas, discuss typical issues that may arise during nonclinical development, and how such issues may be addressed. However, given the complexity of this area, this chapter will only be able to highlight selected aspects that may serve as examples to help understand how the main objectives of nonclinical development can be achieved, the final aim of which is to arrive at a robust risk–benefit assessment to either support the registration of new medications often with a high unmet medical need or to provide the basis for a decision to discontinue programs with an unfavorable safety profile.

TARGET IDENTIFICATION AND VALIDATION

The healthcare community is asking for novel approaches to treat patients, with special emphasis on indications with a high unmet medical need, such as oncology or diseases affecting the central nervous system (CNS) as well as highly complex metabolic or autoimmune diseases. In recent decades, a number of medicines have been successfully developed for many serious conditions, but further improvement has been difficult to achieve and treatment often remains unsatisfactory. Substantial progress often requires truly novel approaches, which includes the need for the identification of new targets to arrive at genuine innovations. In identifying those novel targets the pharmaceutical industry also relies on basic research conducted in academia and published in the literature. The scientific community assumes that results published in a peer-reviewed journal can be taken for granted. However, Prinz et al. [158] reported that validation projects that were started in-house based on published data often could not reproduce published key results. The authors collected data from 67 projects (47 oncology, 12 women's health, and 8 cardiovascular). In only ~20–25% of the projects relevant published data were in line with in-house findings. In almost two-thirds of the projects, there were inconsistencies between published data and in-house data that either considerably prolonged the duration of the target validation process or, in most cases, resulted in termination of the projects because the evidence generated for the therapeutic hypothesis was insufficient to justify further investment into these projects. Similarly, Begley and Ellis [8] were not able to reproduce 47 of 53 landmark publications on basic cancer research published in top journals from

reputable laboratories. Prinz et al. [158] concluded that data on potential drug targets should be viewed with caution, and underline the importance of confirmatory validation studies for pharmaceutical companies and academia before larger investments are made in assay development, high-throughput screening campaigns, lead optimization, and animal testing.

Another important consideration regarding the selection of suitable targets is their specificity in relation to binding molecules, and vice versa. Lounkine et al. [126] have reported on the application of an in silico computational strategy to assess the potential of several 100 marketed drugs regarding their liability to bind to previously unrecognized targets and their possible relevance to the development of known adverse clinical side effects. The aim was to discover unintended secondary ("off") targets through which these drugs may exert adverse drug reactions and to propose tools that may supplement empirical data. Indeed, they identified COX-1 as an off-target for chlorotrianisene, a synthetic estrogen associated with abdominal pain as a side effect. The clinical relevance of this inhibition was demonstrated in whole-human blood platelet aggregation assays. The authors proposed, as a conclusion, that their approach may be useful to predict toxicological liability at the stage of drug discovery. Classical examples for medications that were withdrawn from the market due to off-target toxicity include fen-phen and terfenadine. The former was found to be associated with valvular heart disease (VHD). Rothman et al. [168] concluded that "serotonic medications, which do not activate 5-HT_{2B} receptors, are unlikely to produce VHD." It appears that this liability is brought about by the established role of serotonin as a mitogen [168,177,201], but the specific binding receptor, not surprisingly, mediates the critical effect. In the case of terfenadine, the well-known liability to prolong the QT interval [169] is mediated through its binding to the hERG (human ether-à-go-go related gene) encoded protein K_v11.1, which is the alpha subunit of a potassium ion channel [170] and is also referred to as an hERG channel. In addition, the active metabolite of terfenadine, terfenadine carboxylate, does not show this liability. The case of terfenadine (marketed eg, as Seldane in the United States) led to the development of applicable guidelines (see later) to characterize and manage risk associated with this type of cardiotoxicity. With this molecule, two factors were involved in its cardiotoxicity, ie, binding of the prodrug terfenadine only to the hERG channel as well as an alteration of the metabolic state to increase systemic exposures to the parent drug, terfenadine itself, rather than the active metabolite that does not have this liability [169]. Since then, screening for hERG channel binding has become an integral part of early drug safety assessment, and the field of safety pharmacology has developed into a mature discipline.

Molecules with high target promiscuity or targets with high binding promiscuity or a combination of both intrinsically increase the risk of off-target toxicity at concentrations that may be clinically relevant. Therefore an integrated approach from discovery to the development of new medicines on the market will aid in the selection of new molecules and targets that a priori have a lower liability and therefore may result in more favorable safety profiles.

Biopharmaceuticals, in general, have lower intrinsic liability than small molecules to exert off-target effects and experience to-date indicates that their safety profile will often be dominated by on-target effects also referred to as "exaggerated pharmacology." The intrinsic advantage of this propensity is obvious, as it will allow predicting the occurrence of side effects from exposures at the target, which can be achieved by pharmacodynamic–pharmacokinetic modeling (PK/PD). However, with increasing knowledge about the design of protein scaffolds, there have also been reports about the potential for off-target toxicities for biopharmaceuticals [15,79,82,172], and given the diversity and nature of this class of compounds, it is increasingly acknowledged that biopharmaceuticals need to be carefully evaluated for their potential to cause adverse effects that are not a function of their primary mode of action.

PRINCIPAL ASPECTS OF PRECLINICAL TOXICOLOGY TESTING

Before a new compound can enter clinical trials for the first time the side-effect profile has to be characterized in initial nonclinical safety studies in vitro and in vivo. Those studies are aimed at identifying potential target organs of toxicity, contributing to determining the starting dose in first in human (FIH) trials, and guiding the monitoring program in clinical studies. To support longer clinical studies, the duration of the preclinical studies has to be extended. Nonclinical and clinical development remain closely intertwined from the early beginning until application for a marketing authorization.

Dose Selection for Toxicity Studies with Small Molecules

The first in vivo studies are preliminary dose-range-finding studies in a small number of animals. Based on the results generated applicable doses are selected for the first GLP (good laboratory practice) studies, usually 4-week toxicity studies in rodents and nonrodents. These first studies are intended to characterize the general safety profile including the identification of potential target organs of toxicity and thereby to contribute to setting the starting dose in the first human trial.

A common challenge during this early stage of preclinical development is the availability of sufficient drug substance at the required quality. Limitations in this regard may lead to the conduct of inadequate dose-range finding studies and the selection of doses that fail to show an MTD (maximum tolerated dose) in the subsequent GLP studies. For small molecules, however, it is mandatory to characterize the dose range in its entirety, comprising the no observed adverse effect level (NOAEL) up to the MTD over the duration of the study tested [90]. This principle is key to the characterization of the safety profile of small molecules, although it is not essential to demonstrate the MTD in every study. If no MTD can be demonstrated, a limit dose of 1000 mg/kg/day is usually considered an acceptable high dose in all species provided a mean exposure margin of 10-fold the clinical exposure is achieved. It is of note, however, that suitable high doses are not necessarily based on the same considerations depending on the study type, and specific recommendations are laid out for reproductive and carcinogenicity studies. Limit doses also vary depending on the dose in humans. The various concepts regarding limit doses in the absence of an MTD are summarized in Table 2.1. For biopharmaceuticals, 10-fold multiples compared to human exposures should generally be aimed for ICHS6(R1) [102].

As can be seen from Table 2.1 above, an exposure margin of 50-fold the clinical exposure is considered acceptable as the maximum dose in toxicology studies under the provisions detailed for the US. The obvious limitation is, however, that the estimate of the required clinical exposure may not be correct, particularly during earlier stages of development, and is likely to change once the first pharmacodynamic investigations in humans have been conducted and even later in development, with increasing knowledge regarding the required systemic exposures in humans to achieve efficacy. Actually, maximum recommended human doses (MRHDs) and associated clinical exposures may be a moving target until advanced stages of clinical development. Consequently, safety margins keep changing that complicates dose selection for toxicology studies based on exposures.

In order to avoid repeating toxicity studies and thus wasting cost, time, resources, and animals; doses for the toxicity studies need to be high enough to fulfill regulatory requirements as outlined in the ICH M3(R2) guideline. Clearly, the MTD concept is the most robust approach to guide dose selection and shield against

TABLE 2.1 Overview on Limit Doses Laid Out in ICH Guidance Documents

Guideline	Limit Dose (Human) (mg/person/day)	Limit Dose (Rodent and Nonrodent) (mg/kg/day)	Systemic Exposure Multiples -Fold	Other	Comments
ICH M3(R2)	≤1000	1000	≥10	n/a[a]	
	≥1000	Up to 2000	10	MFD[b]	Whichever is lower
	n/a	2000	<Human exposure	Up to MFD	
	n/a	n/a	50[c]	n/a	To support Phase III clinical trials in the United States, dose-limiting toxicity to be identified in at least one species when using the 50-fold margin as a limit dose.[d]
ICH S1C(R2) carcinogenicity	≤500	1500	≥10	n/a	
	>500	n/a	≥25	Up to MFD	
ICH S5(R2) reproduction	n/a	1000	n/a	n/a	

Note: If genotoxicity endpoints are to be incorporated into a general toxicity study, then an appropriate maximum dose should be selected based on an MFD, MTD or limit dose of 1000 mg/kg/day.

[a] *n/a = not applicable.*

[b] *MFD = maximum feasible dose.*

[c] *According to ICH M3(R2) usually based on group mean AUC values [see Note 1] of the parent drug or the pharmacologically active molecule of a prodrug. […] Note 1: […] "exposure" generally means group mean AUC. In some circumstances (eg, if the compound or compound class is known to produce acute functional cardiovascular changes or central nervous system-related clinical signs) it might be appropriate to base the exposure margin on group mean C_{max} values rather than AUC.*

[d] *If this is not the case, a study of one-month or longer duration in one species that is conducted at the 1000 mg/kg limit dose, MFD or MTD, whichever is lowest, is recommended. However, on a case-by-case basis this study might not be warranted if a study of a shorter duration identifies dose-limiting toxicity at doses higher than those resulting in a 50-fold exposure margin.*

TABLE 2.2 Dose and Duration of General Toxicology Studies to Support Clinical Trials in Nononcology and Oncology Indications

Nononcology [90]	**Oncology [104][a]**
Nonclinical studies of equal or longer duration are needed to support clinical trials of respective length.	Treatment can continue according to the patient's response and can continue beyond the duration of the completed toxicology studies.
Maximal exposure in clinical trials usually limited by exposure in animals.	Highest dose or exposure tested in the nonclinical studies does not limit exposures in cancer patients.
Longer-term toxicology studies (often 3 months) are required to support phase II clinical trials.	Nonclinical data to support Phase I clinical trials are sufficient for moving into Phase II clinical trials.
For clinical trials of >6 months, chronic studies are needed (usually prior to phase III trials).	3-month toxicology studies are needed prior to phase III clinical trials. Further characterization may be done postmarketing.

[a]*Applicable to trials in patients with advanced disease and limited therapeutic options.*

unexpected outcomes later in development, although indeed, for compounds with a benign safety profile, alternative approaches as presented in Table 2.1 may need to be considered. The concept behind dosing up to an MTD is based on the understanding that this approach is most likely to identify all potential target organs of toxicity, but will not result in unspecific toxicity related to excessive systemic exposures. Mortality usually indicates that the MTD has been exceeded, although this may be difficult to establish and is always subject to an integrated assessment of the overall pattern of observations. Dosing below the MTD may not allow for full characterization of all aspects of the safety profile and the results generated therefore may not be suitable to fully support the required clinical monitoring program and consequently may expose patients to unnecessary risks. Failure to establish suitable high-dose levels may thus lead to a hold of clinical programs called for by regulatory authorities until the pivotal preclinical safety studies have been repeated at appropriate exposures.

Another important aspect is the dosing schedule. If the drug is supposed to be given twice or more often/day in the clinic, preclinical dosing schemes should generally mimic the clinical situation, unless exposures in animals can be demonstrated to be adequate following less frequent dosing. Multiple dosing is usually required for drugs with a short half-life. Therefore prior to planning the GLP toxicity studies the pharmacokinetic profile in the respective animal species has to be characterized.

In oncology, special approaches to select doses and treatment schemes in the toxicology studies apply if the first trials are to be conducted in patients with advanced disease and limited therapeutic options [104]. This approach does not relate to drugs that may be tested in healthy volunteers, in which case the same principles have to be followed as laid out above for any drug in development. In oncology, in the clinical setting, a drug may not be given daily, but in cycles. The intended schedule in patients also has to be adopted in the toxicology studies. There is some flexibility in that animals in the toxicology studies can be treated more frequently than patients in the clinical trials. However, in doing so one has to keep in mind that more frequent dosing in toxicology studies may cause more severe toxicity and the drug substance may only be tolerated at lower doses. This will have an impact on the starting dose in the FIH study, which may be lower and hence prolong the dose escalation phase of the clinical trial and unnecessarily expose patients to subtherapeutic doses. As opposed to nonlife-threatening indications, neither the duration nor the dose in advanced stage cancer patients is limited by the duration of the nonclinical studies or the maximal systemic exposures achieved in animals. To illustrate the difference between nononcology and oncology programs, refer to Table 2.2. Subchronic studies are sufficient to support marketing of anticancer medications with the aim to provide access to efficacious treatments in the shortest possible time. However, many oncology drugs are characterized more extensively following marketing, and even carcinogenicity studies may be conducted.

Dose Selection for Toxicity Studies with Biopharmaceuticals

The same general principles as outlined for small molecules also apply to the selection of dose levels investigating the toxicology profile of biopharmaceuticals in nonclinical studies [102]. However, due to their high specificity to the human target, biopharmaceuticals may show little or no toxicity in the animal species used for safety testing and it may not be possible to define a maximum tolerated dose. In these cases, scientific justification of the rationale for the dose selection and projected multiples of human exposure have to be provided [102]. Rather than focusing on the dose-selection aspect of a toxicity study the main focus for biopharmaceuticals is to identify a species in which the molecule is biologically active, ie, a pharmacologically relevant species. The toxicity of most biopharmaceuticals is related to their targeted mechanism of action and typically becomes evident as exaggerated pharmacology, also referred to as "on-target toxicity," whereas the side-effect profile

of small molecules typically is mediated through "off-target toxicity," ie, via binding to other than the intended primary pharmacodynamic target.

As for small molecules, doses selected for biopharmaceuticals need to be justified. The rationale should take the dose–response relationship into account. ICH S6(R1) specifies that PK/PD (pharmacokinetic–pharmacodynamic) approaches (eg, simple exposure–response relationships or more complex modeling and simulation approaches) can assist in high-dose selection by identifying (1) a dose that provides the maximum intended pharmacological effect in the preclinical species; and (2) a dose that provides an approximately 10-fold exposure multiple over the maximum exposure to be achieved in the clinic. The higher of these two doses should be chosen for the high-dose group in preclinical toxicity studies unless there is justification for using a lower dose (eg, maximum feasible dose). In addition, differences in target binding and pharmacological activity should be taken into account to adjust the exposure margin over the highest clinical exposure. However, should no toxicity be evident at the doses selected adopting this approach then higher exposures are considered unlikely to provide any additional useful information.

In practice, many biopharmaceuticals show a benign safety profile even at high multiples of clinical exposure. In general, since many of them are proteins, too high exposures may lead to unspecific effects related to protein overload that does not aid in characterizing the intrinsic safety profile of a biopharmaceutical in development. Furthermore, foreign proteins are prone to elicit an immune response in the host, which can lead to a loss of pharmacological activity through the formation of neutralizing antibodies. Generally, many biopharmaceuticals intended for humans are immunogenic in animals and measurement of antidrug antibodies (ADAs) should be performed in order to aid in the interpretation of these studies both from a toxicological as well as from a pharmacological point of view. The pharmacological response in the toxicological species should preferably be measurable through a pharmacodynamic endpoint that also allows monitoring the strength of the signal in the presence of ADAs to provide information about whether these may be neutralizing the response and if so, to what extent. There is little sense in running toxicity studies longer than until pharmacological activity may have been lost due to the formation of neutralizing antibodies in a large proportion of animals, but typically, the immune response is variable, as in humans. On the other hand, ADAs may not be neutralizing in which case pharmacodynamic endpoints help demonstrate the maintenance of the pharmacological activity and relevance of the species studied. It is of note, however, that the formation of ADAs in preclinical species is species-specific and is not predictive of a potential for antibody formation in humans.

Species Selection for Small Molecules

For small molecules nonclinical toxicity testing has to be conducted in two species, ie, a rodent and a nonrodent [90]. The rodent species is usually the rat and for nonrodents usually dogs, mini-pigs, or monkeys are utilized. For ethical reasons nonhuman primates are only the last resort and whenever possible the dog or increasingly the mini-pig are selected as nonrodent species. In reproductive toxicity studies, typically the rabbit is selected as nonrodent, unless in very special circumstances rendering this model irrelevant.

The preclinical toxicology species should be predictive for humans. This requires that the drug metabolism and pharmacokinetic (DMPK) profiles of a compound are similar in animals and humans. Metabolites that are not formed in the nonclinical test species or are formed in humans at disproportionately higher levels than in any of the animal species during standard toxicology testing may require additional testing in toxicological studies. A major metabolite is considered to be formed in humans at >10% of parent systemic exposure [based on the area under the concentration curve (AUC)] [65]. Therefore it is prudent to conduct a thorough cross-species metabolism profile and select the preclinical species based on their metabolic pathways. The rush into toxicity studies using the default species rat and dog without this knowledge can later be proven to have been the wrong choice resulting in the need to basically start a nonclinical testing program de novo in more appropriate animal models.

Species Selection for Biopharmaceuticals

The development of biopharmaceuticals adds a new layer of complexity. Those compounds have to be tested in at least one pharmacologically active species [102]. Initially, biotechnology-derived pharmaceuticals were developed in the early 1980s, and since then advances in bioengineering have enabled the development of novel efficacious biopharmaceuticals, particularly in areas with a high unmet medical need, such as oncology; the function of these molecules is brought about by the very specific targeting of molecular pathways in humans that cause a particular disease (eg, Ref. [122]). This specificity has great advantages as it can eliminate the potential for toxicity that is not related to the primary mode of action. Consequently, the toxicity of biopharmaceuticals is usually rather more consistent with exaggerated pharmacology than with the off-target toxicity that is typical for small molecules and can often be

predicted based on the understanding of the intended function. However, this specificity to the human target comes with the challenge of identifying a preclinical species for toxicology testing that is pharmacologically relevant. In cases where no pharmacologically active species can be identified the use of homologous proteins can also be used for hazard detection and identification of potential adverse effects. However, such studies are usually not useful in quantitative risk assessment [102]. Another challenge are all those situations where specific targets are only expressed in aberrant tissues that are not present in healthy animals, which are used in nonclinical safety studies.

If the pharmacological effects of a new biopharmaceutical are not similar between the toxicological species and humans the results of the studies conducted can lead to wrong or even dangerous conclusions. An alarming example is TGN1412, a humanized antibody binding to the CD28 protein located on T cells inducing activation of those immune cells.

There are several subtypes of T cells, one being the T effector memory (T_{EM}) cells, another one being the T regulatory (T_{reg}) cells, among others. Activation of T_{EM} cells causes proinflammatory cytokine release, whereas the activation of T_{reg} cells induces antiinflammatory cytokine release, thus suppressing and regulating potential side effects of T_{EM} cells. Imbalances of T_{reg} cells have been related to human autoimmune and vascular inflammatory diseases [35,124] and activation of T_{reg} cells has been shown to be effective in the treatment of autoimmune diseases preclinically [10]. Administration of TGN1412 caused a life-threatening cytokine release syndrome in six healthy male volunteers at the initial dose in the FIH trial. This was not predicted from in vivo preclinical studies in cynomolgus monkeys with TGN1412 and in rats with the homologous antibody (JJ316), and from ex vivo studies exposing human peripheral blood mononuclear cells (PBMC) or diluted blood to TGN1412 [192]. Following the outcome of the first clinical trial with TGN1412 a lot of new knowledge has been gained explaining why the preclinical experiments failed to predict the toxic effects in humans.

Why TGN1412 is bound to human and macaque CD28 with equal affinity but only causes a cytokine release syndrome in humans was investigated by Eastwood et al. [41]. The authors provided convincing evidence that the T_{EM} cells, the cell type that responds with the release of proinflammatory cytokines in humans, does not express the target molecule CD28 in cynomolgus monkeys. During differentiation into T_{EM} cells CD28 expression gets lost only in monkeys, but not in humans. Therefore the cynomolgus monkey was not a biologically active species to establish the safety profile for TGN1412.

The functional equivalent JJ316 did not induce a cytokine release syndrome in rats. Laboratory rats are bred and housed under clean conditions and therefore do not have the chance to accumulate large numbers of T_{EM} cells that can release proinflammatory cytokines [88]. In addition, Mueller et al. [138] found that the activation pattern in rats occurs in two waves. First, there is fast and transient activation of both conventional T_{EM} and T_{reg} cells, followed by a second wave that exclusively activates T_{reg} cells. It is believed that in rats the activation and expansion of T_{reg} cells is so fast that they suppress proinflammatory cytokine release from T_{EM} cells before they reach levels causing clinical symptoms.

Neither the ex vivo cytokine release assays using human peripheral blood mononuclear cells (PBMCs) nor diluted human blood showed cytokine release when exposed to an aqueous solution of TGN1412 [61]. In the meantime it has been discovered that immobilization of the antibody was able to induce the release of proinflammatory cytokines in human PBMCs [184]. Roemer et al. [167] found that when human PBMCs are cultured at high cell density, soluble TGN1412 can subsequently activate the cells. The authors hypothesized that both monocytes and T-cells upregulate functional activity, possibly by acquiring tissue-like properties during high-density culture. Eastwood et al. [42] demonstrated that the severity of the adverse response to TGN1412 correlated with the level of IL-2 release in a solid phase assay.

The experimental conduct of cytokine release assays remains a matter of ongoing debate [74]. There is no formal agreement on assay formats, validation protocols, or appropriate standard procedures on how a cytokine release assay should be conducted and due to the inherent variety of molecule types, a case-by-case approach is needed [108]. A negative in vitro cytokine release assay is still a challenge since the results cannot be exclusively relied on to predict the definitive absence of a respective risk in humans. Unfortunately, in this as well as in many other contexts, absence of evidence is not evidence of absence. Recently issued guidelines on immunogenicity [53,71] address, among other aspects, the need to predict a cytokine response in humans from preclinical in vitro and in vivo animal data, and discuss strategies for testing.

This example emphasizes that knowledge of the nature and comparability of the pharmacological effects in animals and human are of paramount importance in the development of novel biopharmaceuticals.

A general overview of the different aspects that need to be considered when selecting the species and doses for preclinical toxicity studies with either small molecules or biopharmaceuticals is given in Table 2.3.

TABLE 2.3 Criteria for the Selection of Species and Dose for Small Molecules and Biopharmaceuticals

Small molecules	Biopharmaceuticals
SPECIES REQUIRED	
One rodent and one nonrodent species	Pharmacologically relevant species
Selected based on • Comparative in vitro cross-species metabolism data incl. humans	Selected based on • Homology of the target compared to humans • Target-binding affinity • Receptor ligand occupancy • Functional activity (CAVEAT: Binding is not = function!)
HIGH-DOSE SELECTION	
According to ICH M3(R2), S1C(R2) and S5(R2) • MTD[c] (preferred) • Limit dose • Maximum feasible dose • Human systemic exposures and sufficient multiples thereof	Highest dose of either of the two below: • A dose that provides the maximum intended pharmacological effect in the preclinical species • A dose that provides approximately 10-fold exposure multiples over the maximum systemic exposure to be achieved in the clinic
Toxicity driven by unknown endpoints	Pharmacology driven by known endpoints
CONCEPTS TO DERIVE STARTING DOSE IN FIH[A] TRIALS	
NOAEL[d]-driven	MABEL[b] or PAD[e] driven
Maximum recommended starting dose • Based on toxicity • Determination of human equivalent dose based on body surface area • Application of safety factor	• Based on pharmacology • Need of pharmacologically relevant assays in humans and animals (in vitro and in vivo) • Pharmacokinetic/pharmacodynamic modeling • Adjustment for interspecies differences in affinity and potency
Highest dose thought to be safe	Lowest dose thought to be active

[a] *FIH = First in human.*
[b] *MABEL = Minimal anticipated biological effect level [48].*
[c] *MTD = Maximum tolerated dose [90].*
[d] *NOAEL = No observed adverse effect level [63,90].*
[e] *PAD = Pharmacologically active dose [63].*

PHASE I

Following years of nonclinical work during drug discovery, lead optimization, animal testing, pharmacokinetic characterization, and toxicology investigations the administration of the first dose of a novel compound in humans is an exciting step. The entry into clinical development is designated as Phase I. It is the first step to determine if the predictions made from preclinical models will also translate into the clinic. In most cases, healthy volunteers are tested. In some instances, the FIH study may be conducted in patients, such as in oncology. Historically, drugs to treat cancer patients have been very toxic themselves, prohibiting the administration of those test materials to healthy volunteers, and in this disease area FIH studies are regularly conducted in patients. Newer anticancer medicines have fewer side effects, opening the possibility to also include healthy volunteers in the clinical development process. The dose in the FIH study is well below a dose that caused toxicity in animals and the first dose is usually uneventful. However, this cannot be taken for granted.

TGN1412

Sometimes the unexpected can happen and the events on March 13, 2006 resulting from the administration of TGN1412 to six healthy volunteers are a warning example. TGN1412 is a humanized IgG4 agonistic anti-CD28 monoclonal antibody designed to stimulate T cells by activating CD28 signaling without the need for prior activation of the T-cell antigen receptor. Due to this ability TGN1412 was also called a "superagonist" [192]. It was intended for the treatment of hematological malignancies, such as B-cell chronic lymphatic leukaemia (B-CLL) and autoimmune/inflammatory diseases, such as rheumathoid arthritis (RA).

In the FIH trial, all six healthy male volunteers developed a cytokine release syndrome with multiorgan failure requiring intensive treatment and supportive care by the intensive care unit [61]. FIH trials until then had a very good safety record and, as far as the Expert Scientific Group could determine, the TGN1412 trial outcome was unprecedented. TGN1412 underwent preclinical testing according to current regulatory requirements. TGN1412 cross-reacted with CD28 expressed on T cells

from humans and cynomolgus and rhesus monkeys, thus establishing the cynomolgus monkey as an appropriate species for toxicological testing. The compound was well tolerated in cynomolgus monkeys at doses up to 50 mg/kg/week for 4 consecutive weeks, and this dose level was designated to be the NOAEL. Moderate elevations of IL-2, IL-5, and IL-6 serum levels were observed in individual animals, but no clinical signs of a cytokine release syndrome were observed. The FIH dose was derived using the FDA guidance [63] to calculate a human-equivalent dose. After applying a default factor of 10 the maximum recommended starting dose in healthy volunteers was estimated to be 1.6 mg/kg. The company then applied an additional safety margin and proposed a starting dose of 0.1 mg/kg [192]. This dose was 500-fold below the NOAEL established in cynomolgus monkeys. Despite meeting all regulatory requirements this approach failed to establish a safe starting dose in the FIH trial with TGN1412. However, it should be noted that when calculating the human starting dose for TGN1412 no consideration had been given to the pharmacologically active dose (PAD). In step 5 of the FDA guidance document it is noted that for certain classes of drugs or biopharmaceuticals like monoclonal antibodies toxicity may arise from exaggerated pharmacologic effects. In such a case, the PAD may be a more sensitive indicator of potential toxicity than the NOAEL. However, the FDA guidance document focuses on the NOAEL approach and does not give much detail on the PAD because "selection of a PAD depends on many factors and differs markedly among pharmacologically drug classes and clinical indications; therefore, selection of a PAD is beyond the scope of this guidance."

Regulatory requirements laid out in applicable guidelines are only guidance documents that assist in the general principles and scientific standards that should be met. They cannot cover all possibilities and in no way can they be used as a check box system. Sponsors are the experts on their investigational medicinal products and have the responsibility to conduct a thorough preclinical evaluation and a critical review of the available data. Of course, "hindsight is 20/20" and with all we know today the decision to treat six healthy volunteers with 0.1 mg/kg TGN1412 within 10 min would not have been made. But could the disaster have been avoided based on the knowledge available prior to trial initiation?

Based on pharmacological effects in healthy and arthritic rats using the rat CD28-specific homologous antibody JJ316 and pharmacological effects of TGN1412 in rhesus and cynomolgus monkeys the minimal anticipated biological effect level (MABEL) could be considered to be between 0.1 and 1 mg/kg [61]. The ESG calculated a safe starting dose in humans of 5 μg/kg, considering the MABEL dose to be 0.5 mg/kg and applying a safety factor of 1/100th as proposed for microdosing in the respective European Medicines Agency (EMA) [44] position paper. The association of the British Pharmaceutical Industry and the BioIndustry Association [139] calculated that the initial FIH dose of 0.1 mg/kg resulted in >90% receptor occupancy and hence was likely to achieve the maximum pharmacological effect. For a drug like TGN1412 with a novel agonistic mode of action, an initial receptor occupancy of below 10% may be more appropriate. This level of occupancy was predicted to be achieved with a dose of 1 μg/kg, ie, a dose by a factor of 100 lower than the actual starting dose, whereas a dose of 5 μg/kg would have resulted in approximately 33% receptor occupancy [87,127,140].

TGN1412 is an excellent example of how the dose makes the poison, an observation originally made more than 500 years ago by Paracelsus. TGN1412 is now called TAB08 and has been safely administered to healthy volunteers at doses of 0.1–7 μg/kg [190]. The starting dose was 1000-fold less than applied in the 2006 clinical trial. At 5 and 7 μg/kg (15–20 times less than the dose used in the 2006 trial) evidence showed that TAB08 had stimulated an antiinflammatory response in the absence of cytokine release. Clinical development of TAB08 continues in RA patients [193].

PHASE II

During this phase of clinical development the drug is given to patients to further assess efficacy and safety with the aim to establish the basis for the pivotal phase III trials. In the nonclinical arena additional toxicity studies, mostly repeated dose toxicity studies of longer duration and reproductive toxicity studies are conducted. The CMC (chemistry, manufacturing, and control) aspects of drug development are equally complex and closely intertwined with the nonclinical and clinical fields, and changes affecting CMC often have a knock-on effect on the latter areas.

Introduction of Salt or Change of Salt Form

A "new drug substance" is defined as follows in ICH Q3A(R2) [93]: "The designated therapeutic moiety that has not been previously registered in a region or member state (also referred to as a new molecular entity or new chemical entity). It can be a complex, simple ester, or salt of a previously approved substance." This implies that a change in salt form may render the available results from earlier safety studies with another form of the drug substance at least partly invalid to support registration, if the new salt form is considered to represent a new drug substance.

Salts are used to alter the physical or chemical properties of a drug, but are not intended to change the intrinsic properties of the therapeutic moiety. Changing the

salt form can improve solubility and thereby enhance absorption and increase systemic bioavailability. Salts can also improve stability and therefore prolong the shelf life of a drug; furthermore, the formulation of the final product is influenced by the salt form. It is estimated that about 50% of the drugs on the market are administered as salts [7]. A formal salt selection program takes time and requires compound. In the race to bring new compounds early into clinical development the toxicology and FIH studies may be conducted with a suboptimal salt or with the free acid or base of the drug. The later in development the final salt form is introduced the more studies may need to be repeated. Ideally a change in salt form should only take place prior to initiating long-term toxicity studies.

Since a new salt form can change the bioavailability of a drug pharmacokinetic bridging studies need to be conducted in order to show either bioequivalence, ie, comparable systemic exposures and DMPK profiles, or get information on how to adjust doses for future human or toxicology studies. For drugs with a narrow therapeutic window, the change in salt form can have a significant influence on the dose required to achieve bioequivalence. In some cases, toxicological bridging studies may be needed in addition prior to the conduct of longer term studies.

Impurities

During drug development the compound manufacturing process continues to be optimized and until final procedures are definitively established the impurity profile of an active pharmaceutical ingredient (API, also referred to as drug substance or therapeutic moiety) or of the final formulation of the API (referred to as drug product) can change. Modifications of the impurity profile can, for example, be caused by changes in the route of synthesis, in starting materials or intermediates or it could be a degradation product developing over time. Changes can still occur postmarketing, in the event of further CMC modifications, such as for generic products.

A number of guidelines has been issued to regulate maximum amounts of impurities allowed in drug substances and drug products, and also to specifically address genotoxic impurities, including:

- ICH Q3A(R2) Impurities in New Drug Substances [93]
- ICH Q3B(R2) Impurities in New Drug Products [94]
- ICH Q3C(R5) Guideline for Residual Solvents [95]
- ICH Q3D Guideline for elemental impurities [96]
- ICH M7, Assessment and Control of DNA Reactive (Mutagenic) Impurities in Pharmaceuticals to Limit Potential Carcinogenic Risk [91]
- ICH M7(R1) Draft Addendum – Assessment and Control of DNA Reactive (Mutagenic) Impurities in Pharmaceuticals to Limit Potential Carcinogenic Risk. Application of the Principles of the ICH M7 Guideline to Calculation of Compound-Specific Acceptable Intakes [92]
- EMA, Guideline on the Limits of Genotoxic Impurities [47]
- EMA, Questions and Answers on the CHMP Guideline on the limits of genotoxic impurities [51]
- FDA Draft Guidance for Industry: Genotoxic and Carcinogenic Impurities in Drug Substances and Products: Recommended Approaches [66]

These regulations are unfortunately not fully consistent with each other and are not always clear with respect to their specific recommendations. Importantly, in general only selected aspects are considered and addressed. The impurity limits laid out in the ICH Q3x guidelines issued several years before ICH M7 may be well above those laid out in the recent ICH M7 guideline, regardless of any genotoxic potential (Note 1 in ICH M7). On the other hand, the focus of ICH M7 is on mutagenic (but not clastogenic or aneugenic) genotoxic mechanisms, which can be tested in the bacterial reverse mutation assay. In contrast to ICH Q3x, ICH M7 does not take daily doses into consideration, whereas the duration of dosing is an integral part of concepts laid out in both ICH Q3x as well as in ICH M7. ICH Q3A(R2) and ICH Q3B(R2) define so-called reporting, identification, and qualification thresholds for impurities in the drug substance and drug product, which reflect quantitative (relative or absolute) limits of respective impurity in the drug substance or product above which applicable action has to be taken. If the qualification threshold is exceeded, a toxicological assessment must be made based on either published information or on available or even new safety studies. Any impurity that was adequately tested in safety studies or an impurity that is a significant metabolite in animal studies can be considered qualified. However, if new impurities arise after the toxicology program has been finalized additional studies may be needed. If no information is available, as a minimum, in vitro genotoxicity studies (a study investigating point mutations and a chromosomal aberration test) and a repeated dose toxicity study in one species are required for qualification. The duration of the repeated dose toxicity study can vary from 14 to 90 days to support short to chronic treatment, respectively. Furthermore, ICH Q3x guidelines suggest evaluating "other specific toxicity endpoints, as appropriate," which implies that a scientific judgment is required to review the extent of the investigations considered mandatory. This implies a case-by-case approach and no fixed limits are proposed for such event, but it is implied that an integrated toxicological assessment be made to deduce an applicable safe level based on specific aspects and not using a default approach.

The most problematic impurities are those that are mutagenic, since there is increased concern regarding potential carcinogenicity without a threshold level, although this is a matter of debate and currently under review [92]. Other types of genotoxicants typically have a threshold mechanism and usually do not pose a carcinogenic risk in humans at the levels present as impurities [92]. ICH M7 defines five classes of impurities with respect to their mutagenic and carcinogenic potential; chemicals in class 1 are those that are classified as known mutagenic carcinogens. The general principles described in ICH M7 adopt a threshold of toxicological concern (TTC) concept that was developed to define an acceptable intake for any unstudied chemical that poses a negligible risk of carcinogenicity or other toxic effects. The methods on which the TTC is based are generally considered to be very conservative since they involve a simple linear extrapolation from the dose giving a 50% tumor incidence (TD_{50}) to a 1 in 10^6 incidence, using TD_{50} data for the most sensitive species and most sensitive site of tumor induction. For application of a TTC in the assessment of acceptable limits of mutagenic impurities in drug substances and drug products, a value of 1.5 μg/day corresponding to a theoretical 10^{-5} excess lifetime risk of cancer, can be justified. Some structural groups were identified to be of such high potency that intakes even below the TTC would theoretically be associated with potential for significant carcinogenic risk. This group of high-potency mutagenic carcinogens, referred to as the "cohort of concern," comprises aflatoxin-like-, *N*-nitroso-, and alkyl-azoxy compounds.

ICH M7 also provides specific guidance for applicable higher limits during clinical development, where the duration of treatment is shorter. The concept of ICH M7 is taken further in the recently issued draft addendum ICH M7(R1) that describes appropriate approaches for selected chemicals that are considered to be carcinogens with a likely mutagenic mode of action and gives insight into the applicability of setting threshold levels rather than adopting a linear extrapolation approach for some class 1 chemicals, ie, mutagenic carcinogens.

For some areas of toxicology, there is no specific recommendation in any of these guidelines, such as with respect to the need (or lack) of addressing a potential for reproductive toxicity.

Taken together, all guidelines have to be considered in context and with respect to the target population, dose, and duration of clinical use. The most stringent approach should be chosen for chronic treatment.

In essence, changes in CMC processes during drug development require an integrated approach between disciplines to proactively address potential issues, but still here, unexpected situations may result in challenging situations. Such situations could be those where the formation of a toxic impurity may not have been expected and the impurity has to be reduced to levels that are not technically achievable. In turn, such outcomes my trigger the need for altering the CMC process and/or require additional toxicology studies, which obviously has an impact on the resources needed and may well delay approval of new drug substances at a stage where the majority of studies has been completed.

PHASE III

During this phase of clinical development the drug is given to a sufficient number of patients to gather pivotal information on efficacy and safety; in many indications this may involve several 100 or even 1000 volunteers. Those trials are usually randomized and double-blinded where neither the investigator nor the patient knows if the new therapy or a comparator (ie, placebo or another therapy) is given. These pivotal phase III trials provide the basis for the definitive risk–benefit assessment prior to marketing application and support the registration of a new therapeutic. In the nonclinical arena usually the reproductive toxicology package is completed and the carcinogenicity studies conducted. As in all stages of development, further mechanistic toxicity studies may be performed to support hypotheses about the relevance of nonclinical findings for human safety.

Clinical Hold

Once a compound has entered clinical development testing in animals continues in parallel. In most therapeutic indications the treatment duration in clinical trials must not exceed the treatment duration in nonclinical toxicology studies. There are both predefined points in time at which data are typically compiled by sponsors for discussion with regulators to approve the next phase of development, such as before entering first clinical studies, or at the end of phase II of clinical development, ie, prior to starting the big and expensive pivotal clinical studies, as well as situations where the sponsor has to notify regulators of side effects in specific situations in a predefined short timeframe. In the United States, for example, the sponsor is required to notify the FDA and all participating investigators in an IND (investigational new drug) safety report (ie, 7- or 15-day expedited report) of *potentially serious risks* from clinical trials or *any other source* as soon as possible, but no later than 15 calendar days after the sponsor receives the safety information and determines that the information qualifies for reporting [70]. This definition also embraces findings from toxicology studies and implies that the responsibility of judgment lies with the sponsor. The shorter timeline relates to unexpected fatal or life-threatening adverse reactions.

The agency can order a clinical hold following a review of data at predefined time-points as well as after reporting potentially serious risks, in which case all or some of the investigations conducted under an IND application may be suspended and subjects enrolled in clinical trials may no longer be given the investigational drug. The agency may ask for additional studies to investigate the issue identified. The clinical hold may be lifted if additional animal data can be provided that demonstrate the safe use in the proposed clinical trial. Similar processes and timelines are in place in other regions, such as the EU and are established to ensure patient safety.

Generally, the earlier the stage of development the more weight that is attributed to nonclinical observations. In view of the impact of a clinical hold on the one hand, but unexpected side effects in humans even at advanced stages of development or during postmarketing, which were not anticipated from nonclinical results, the question arises as to how predictive in general preclinical data are for humans. Classical observations that present challenges in development and specific fields of particular concern are discussed in the following.

PREDICTIVITY OF TOXICOLOGICAL FINDINGS FOR HUMAN SAFETY

Typical Issues and How to Deal with Them: Clinical Intolerance, Liver Toxicity, Nervous System and Retinal Toxicity, Endocrine Disorders, Phospholipidosis

Typical issues encountered in preclinical studies include clinical and target organ toxicity in the toxicological species often associated with low safety ratios (SRs), although in general, a safety margin of at least 10 is a minimum requirement (see Table 2.1). SRs are derived from a comparison of systemic drug exposures in patients at therapeutic doses and animals at the NOAEL. For small molecules, SRs would be preferred to be >20, and greater margins are certainly a safeguard particularly in situations where there is a steep dose–response, but for some classes including CNS drugs, SRs often can be <10 and may even be <1. In the latter class, dose-limiting clinical intolerance in animals and healthy volunteers at doses below those tolerated in patients is not uncommon. Typical features may include the lack of a histopathological correlate, reversibility on cessation of dosing, and not infrequently, an amelioration with continued dosing. CNS toxicity generally presents as signs consistent with exaggerated pharmacology, such as tremors, increased or decreased activity/sedation, recumbency, loss of balance/ataxia, hypothermia (rats), seizures/convulsions, and death. Examples include clozapine, haloperidol, bupropion, tricyclic antidepressants (eg, trimiparmin, nortriptyline), benzodiazapines (eg, diazepam) (NIH–TOXNET [142–147]), risperidone [166], and AChE (acetylcholine esterase inhibitors) inhibitors (eg, rivastigmine) [45,60].

Target organs of toxicity for a variety of drugs may include the liver, CNS/PNS (peripheral nervous system), endocrine system, lung or retina, or may feature as phospholipidosis across a number of organs. Hepatotoxicity can be present in one or more preclinical species and generally is predictive for humans. Characteristics may include elevated serum enzymes, increased liver weight, and morphological alterations (such as hepatocyte hypertrophy, vacuolation, lipid deposition, degeneration, and necrosis or hepatobiliary changes). Hepatocyte hypertrophy is often adaptive due to stimulation of drug metabolism and nonadverse, but this change could lead to potentially severe toxicity at higher doses or on prolonged treatment. In contrast, idiosyncratic liver toxicity in man is not predicted from animal studies and often due to metabolic differences in (individual) humans or may be immunologically mediated, which results in higher susceptibility of the individual affected. In general, animal species are poor predictors of adverse human immunological issues [54,130,136,154,156].

Morphological changes of the nervous system are variable and can include findings, such as vacuoles in the neurones, in their axons, in glial cells, and/or in the myelin sheath, as neuronal pigmentation and as necrosis, reflecting neuronal damage. They may be the result of direct neurotoxic action of a drug and/or result from vascular injury. Such alterations may or may not be reversible and/or be associated with a functional deficit. Examples from animal studies include a number of drugs, eg, interacting with the NMDA receptor, such as phencyclidine, MK-801, or memantine [5,29,148,182]. Morphological findings in the CNS are nonmonitorable in the clinic unless they were reliably identifiable by a biomarker indicating a fully reversible functional stage well preceding any changes at the histopathological level. For obvious reasons, such monitoring is severely hampered by medical and technical limitations, and mostly, compounds with such findings are not developed further.

Endocrine disorders can for instance be caused by dopamine D_2 antagonists through elevated circulating prolactin levels, possibly associated with pituitary and mammary proliferative changes or disruption of male and female reproductive function, or dopamine D_2 agonists that may reduce prolactin levels. Examples for compounds associated with such hormonal alterations include risperidone [165,166], aripiprazole [1,2], and bromocriptine [153]. In the rat, pregnancy loss due to systemic (maternal) hypoprolactinemia is a known effect since pregnancy is established and maintained by prolactin rather than progesterone in this species unlike

in rabbits and humans. The early phase of pregnancy in rats is dominated by surges of prolactin produced by the maternal anterior pituitary. Therefore any effects on early pregnancy due to reduced maternal systemic prolactin levels are not predictive for humans but could hamper the validity of reproductive studies in the rat depending on the type and severity of effects secondary to a potential hypoprolactinemia. Furthermore, such effects could potentially confound the interpretation of animal studies [69,181].

Retinal atrophy particularly in the albino rat may feature as a loss of nuclei of the outer nuclear layer with thinning of the photoreceptor layer and progress to the loss of all layers with disruption of the pigment epithelial layer. This alteration has been described for a number of drugs, eg, pregabalin [128], pramipexole [135], aripiprazole [1,2], and citalopram [19], none of which has been associated with retinal changes in humans, suggesting a limitation regarding the predictive value of this type of toxicological finding in albino rats for human safety, although this cannot be taken for granted, and respective measures for each project have to be taken to address this issue in the event of similar observations.

Phospholipidosis caused by cationic amphiphilic drugs for many different indications is characterized by excessive accumulation of phospholipids in cells, usually within lysosomes, and presents with a lamellar structure often in lungs and liver but possibly also in lymphoid and other tissues (eg, kidney). Different species and even strains within species and also different age groups may not react similarly to the same agent, and overall, the response to a specific cationic amphiphilic drug in a particular species is considered unpredictable [137]. The severity (extent of accumulation) varies between drugs. Phospholipidosis may reflect an adaptive rather than toxic response and does not usually disrupt organ function. There are no validated biomarkers available for clinical use as yet. Examples include amiodarone, imipramine, and fluoxetine [137,141,152,161,171]. The significance of phospholipidosis in preclinical animal studies for human risk assessment is a matter of ongoing debate [68,162].

If issues are identified during nonclinical development, the following steps are recommended:

1. Do not stop development immediately but:
2. Review the finding in detail first to answer the following questions:
   - **a.** Is it a real observation or could it be an artifact?
   - **b.** What is its nature?
   - **c.** Is it an exacerbation of a spontaneous finding?
   - **d.** Is it a known class finding?
   - **e.** Was the finding statistically significant or does it only affect individual animals; if so are these individuals representative for the group, specifically susceptible or outliers?
   - **f.** Could it be a chance finding?
   - **g.** May the finding be species-specific?
   - **h.** Is the observation reversible?
   - **i.** Does the observation deteriorate with ongoing treatment—perhaps to an irreversible stage?
   - **j.** What is the degree of severity?
   - **k.** Can the observation be reliably monitored in the clinic?
   - **l.** Is the finding considered predictive or relevant for man?
   - **m.** Can this question be answered at all at the respective stage of development?
   - **n.** What are the predicted safety ratios?
   - **o.** Is the sensitivity comparable between species?
   - **p.** Can this question be answered at all?
   - **q.** If not, are the safety ratios a reliable tool to estimate human risk, or do additional factors need to be taken into account?
3. The answers to these questions will inform on the overall risk for further development of the project and, if deemed appropriate, provide a sound basis for working out an appropriate action plan. It is recommended to proactively enter into a dialogue for scientific advice with governmental regulators at an early stage to establish whether the action plan is deemed appropriate and/or how it might need to be modified for a successful testing strategy.

Preclinical issues and low SRs are not necessarily impediments to successful drug development. Many issues would be "stoppers" for new drugs in "soft" indications but not necessarily for indications with a high unmet medical need. Some preclinical issues do not appear to be predictive for patients. Others are predictive but are monitorable clinically and safety can be ensured, whereas nonmonitorable and severe toxicities may indeed require discontinuation of further development of the drug concerned.

Cardiotoxicity

The heart is a remarkable organ. With about 100,000 beats a day it pumps approximately 7000 L of blood through a network of vessels that when laid out end-to-end would circle more than twice the planet Earth (97,000 km). It is not a surprise that this vital organ is also prone to toxic insults. There are two major classes of myocardial injury, ie, structural and nonstructural injuries. Cardiotoxicity can be caused by alterations in biochemical pathways, energy metabolism, cellular structures, electrophysiology, and contractility leading to decreased cardiac output and peripheral tissue hypoperfusion. In vitro (eg, ion-channel function, Purkinje fiber assay) and in vivo studies (eg, telemetry, electrocardiography, histopathology) are conducted during the nonclinical

development to investigate potential side effects on the heart. Any finding is of concern and requires an appropriate risk assessment and possibly follow-up studies.

QT Prolongation

During the 1990s several drugs from different therapeutic indications were removed from the market due to drug-induced cardiac arrhythmias (eg, terfenadine (antihistamine), grepafloxacine (antibiotic), sertindole (antipsychotic), cisapride (heartburn)). The overall frequency of those serious adverse events leading eventually to market withdrawal can be extremely low (eg, less than 1 in 100,000 patients experienced TdP with terfenadine) [121]. It was found that those arrhythmias were associated with prolongation of the QTc interval, which may lead to polymorphic ventricular tachycardia, also known as torsade de pointes (TdP), which can be fatal. This led to the development of preclinical [103] and clinical [89] guidance documents requiring drug developers to test for drug-induced QT prolongation prior to seeking drug approval. Since the most common mechanism for QT-interval prolongation by pharmaceuticals is inhibition of the delayed rectifier potassium channel pharmaceutical companies have started early screening for hERG channel-blocking properties. However, not every compound that blocks the hERG channel also induces QT prolongation and possibly the feared TdP. A screening solely toward unwanted hERG effects may sort out promising candidates for drug development. Drugs that block the hERG channel may not cause QT prolongation if they counteract the potential hERG channel block by simultaneous blockage of L-type Ca^{2+} channels, eg, Verapamil [14]. For a new compound that blocks the hERG channel but does not induce QT prolongation in the in vivo telemetry study mixed channel activities may be suspected that does not always preclude further drug development. CIPA (comprehensive in vitro proarrhythmia assay) is a new ILSI (International Life Sciences Institute) initiative that is evaluating the current paradigm of testing and is in the process of proposing a suite of preclinical in vitro and in vivo studies that may sufficiently support clinical development and eliminate the need for a thorough QT/QTc study in the clinic (ICH E14).

Cardiomyopathy

This is a common background lesion in rat toxicology studies [20]. If a dose-dependent increase in incidence and severity is seen in toxicology studies its relevance to humans has to be evaluated. This finding is frequently observed with immunosuppressive compounds. In a study investigating the mechanism of sirolimus-induced myocardial degeneration the finding could be attributed to the activation of latent parvovirus in the hearts of immunosuppressed rats. Subsequently, this effect was not considered to be adverse [46]. In the sirolimus scientific discussion it is also noted that cyclosporin A and tacrolimus have also been reported to induce myocardial degeneration in rats.

The Cardiovascular Safety of Anticancer Therapies

The cardiotoxic potential of cytotoxic chemotherapeutics (eg, anthracyclines) is well known. The so-called "targeted" therapies, which interact with targets that are overexpressed and/or mutated in tumor cells, specifically the protein kinase inhibitors (eg, Gleevec/Glivec, imatinib), have revolutionized the treatment of certain cancers with better tolerability than conventional chemotherapies. However, these kinases are also expressed in cardiac tissue and play a crucial role in normal homeostasis. Consequently, a number of protein kinase inhibitors has been associated with cardiotoxicity in humans [21,151]. There are, however, distinct mechanistic differences in the manifestation of chemotherapy-induced cardiotoxicity. Type I cardiotoxicity causes myocardial damage that is characterized by direct myocyte injury (eg, vacuolation, myofibril disarray, necrosis) resulting in dose-dependent permanent toxicity [57,59,205]. Examples of type I agents are anthracyclines (eg, doxorubicin, daunorubicin, epirubicin), mitoxantrone, and cyclophosphamide [58,59], which are classical cytotoxic anticancer therapies. Type II cardiotoxicity is characterized by myocyte dysfunction that is not dose-related or associated with structural damage and often reversible and therefore has a more favorable prognosis [57]. Examples of type II agents are trastuzumab, sunitinib, imatinib, and lapatinib [58,59], which belong to the new "targeted" cancer therapies.

Cardiovascular side effects have been seen with both small molecules and biopharmaceuticals. Cancer is often treated with a combination of medicines to improve efficacy, but this approach comes with the risk of additive or synergistic side effects. Trastuzumab (Herceptin [85]), a humanized monoclonal antibody approved for the treatment of HER2-overexpressing breast cancer and HER2-overexpressing metastatic gastric or gastroesophageal junction adenocarcinoma, has a boxed warning for cardiomyopathy (Herceptin prescribing information). Congestive heart failure (CHF) occurred in 7% of patients treated with either Herceptin alone or with a combination of anthracycline and cyclophosphamide. However, when all three agents were coadministered the incidence of CHF increased to 28%. A less pronounced synergistic effect (11%) was seen when Herceptin was combined with paclitaxel (Herceptin prescribing information). Although the exact mechanism of Herceptin-induced cardiotoxicity is not fully understood there is evidence that the ErbB2 receptor (synonym for HER2/neu) is involved in growth and survival pathways of adult cardiomyocytes that are probably essential for cell repair [30,36,75,111].

Force and Kolaja [76] reviewed the cardiotoxicity of kinase inhibitors, which can be associated both with the primary ("on-target") as well as with unintended ("off") targets, and the predictivity and translation of preclinical models to clinical outcomes, which had limited success to date. Therefore in clinical practice patient monitoring is crucial in the management of side effects of targeted anticancer therapies [40]. Cancer patients in many indications today have prolonged life expectancy and improved survival rates and the long-term safety of anticancer therapies has to be revealed. Cardiovascular side effects in this medical field have given birth to the emerging clinical discipline of cardio-oncology.

Genotoxicity

Genotoxicity tests are designed to detect compounds that induce genetic damage and are mainly used for the prediction of carcinogenicity [100]. Carcinogenicity is a complicated multistep process, and in experimental animal testing the "gold standard" for human risk assessment still is the 2-year carcinogenicity bioassay in rodents, although ICH is currently reviewing the need and extent of requirements for carcinogenicity testing. The available genotoxicity tests are fairly simple, short-term in vitro and in vivo tests and it is not surprising that their predictivity toward the gold standard is far from perfect. In order to develop safe compounds those tests should show high sensitivity (ie, correctly predict a positive response in the 2-year carcinogenicity assay). But on the other hand, those tests have also been shown to have low specificity (to correctly identify a negative response in the 2-year carcinogenicity assay). This leads to a high number of false-positive assays that need to be evaluated further. This problem has been described in several review articles [34,114–116,196,197] and has led to the development of a revised guideline on genotoxicity testing [100].

Positive Ames Test—What Next?

The false-positive rate of the Ames mutagenicity test is very low. Extensive reviews have shown that many compounds that are mutagenic in the bacterial reverse mutation (Ames) test are indeed rodent carcinogens [132,209]. Due to the established strong correlation between a positive Ames test and a positive rodent carcinogenicity study a positive Ames test requires extensive follow-up testing to assess the mutagenic and carcinogenic potential of the compound. However, there are situations where the Ames test can be false-positive. Compounds that can release amino acids (histidine or tryptophan) into the culture medium can create a false-positive effect [3]. This mode of action has to be kept in mind when testing compounds that are derived from biological material (eg, proteins, peptides, food additives, cosmetics, herbal extracts). Bacterial-specific metabolic activation, such as nitroreductases has been linked to the creation of genotoxic impurities. Therefore positive Ames tests with aromatic nitro compounds may not be predictive for genotoxicity using mammalian assays [117]. AMP397, a novel antiepileptic drug with an aromatic nitro group, was positive in the Ames test, but negative in nitroreductase-deficient Ames tester strains [185]. In addition, no genotoxic activity was determined with AMP397 in several in vivo assays, including a comet assay in the jejunum a tissue where nitroreductases would be present. This example shows the kind of scrutiny and follow-up investigations that may be needed to successfully continue development of an Ames-positive compound even in a nonlife-threatening indication.

Positive In Vitro Mammalian Cell Assay—What Next?

The false-positive rate of the in vitro mammalian cell assay is quite high. Therefore the ICH S2(R1) [100] gives guidance on evaluation of test results and on follow-up test strategies. Those include but are not limited to the assessment of reproducibility, biological significance (statistically significant findings that are still within the historical control range), nonphysiological conditions (pH, osmolality, precipitates), and the concentration-effect relationship (positive only at the highest, most toxic concentration). A positive in vitro genotoxicity test has to be followed by mechanistic information that contributes to the weight of evidence for a lack of relevant genotoxicity. This can include in vitro or in vivo assays, depending on the kind of findings observed. Aneugens affect cell division by interaction with the spindle apparatus and not directly by interacting with DNA. For this mechanism, it might be possible to determine a threshold exposure below which the loss of chromosomes does not occur. Such a compound could be safely given to humans if an appropriate safety margin exists. Clastogens damage chromosomes and if an in vitro test is positive, two negative assays measuring the same endpoints are required in vivo to demonstrate the lack of relevance of the in vitro assay.

Carcinogenicity

Drugs that are indended to be used continuously for at least 6 months have to be tested for their potential to induce tumors. For pharmaceuticals used frequently in an intermittent manner in the treatment of chronic or recurrent conditions, carcinogenicity studies are generally needed. Examples include drugs for the treatment of allergic rhinitis, depression, or anxiety [98]. Carcinogenicity studies are conducted in two rodent species, usually the rat and the mouse [99], although ICH is currently reviewing the need and extent of requirements for carcinogenicity testing [97].

The 2-year bioassay conducted in either species or a combination of a 2-year bioassay in rats with a 6-month transgenic mouse assay in mice are common strategies.

Given the high cost and extensive use of animals those studies are among the last to be conducted during preclinical testing prior to applying for marketing approval. In certain cases, eg, indications of high medical needs, carcinogenicity studies may also be conducted postapproval. Only in exceptional situations, ie, in case of significant cause for concern, carcinogenicity studies may need to be submitted to support clinical trials [90].

Positive Results in Rodent Carcinogenicity Study—What Next?

A positive carcinogenicity study does not necessarily mean the end of development. Contrera et al. [27] reviewed 282 (229 marketed) human pharmaceuticals in the FDA database and found that 44.3% of the compounds had positive carcinogenicity findings. Similarly, Van Oosterhout et al. [202] reported that for nearly 50% of the compounds for which a marketing authorization was applied in Germany and the Netherlands a positive carcinogenicity study was submitted, with the rat being more sensitive than the mouse. Once a positive finding is discovered its relevance to humans has to be determined. Genotoxic compounds are usually sorted out early in development and will not make it to the stage of carcinogenicity testing. One exception is the development of drugs to treat cancer or other life-threatening conditions where the benefit outweighs the risk of possibly developing a drug-induced tumor. Several authors have critically reviewed the relevance of the 2-year carcinogenicity assay for human risk assessment and some also proposed alternative testing strategies [6,163,180]. A review of these alternative testing strategies is outside the scope of this chapter, which focuses on how to deal with a positive finding in carcinogenicity studies for nongenotoxic compounds and how to assess the human risk.

In the event of a positive carcinogenicity study, the principal initial approach is to first evaluate a mode of action (MOA) of tumorigenesis in animals and then to assess its relevance for humans. A framework for analyzing the MOA by which chemicals induce tumors in laboratory animals has been developed by the International Programme on Chemical Safety (IPCS) and was published by Sonich-Mullin et al. [183]. According to these authors, the MOA analysis includes the following steps:

1. Introduction: The description of the cancer endpoint/endpoints.
2. Postulated mode of action (theory of the case): The description of the sequence of events on the path to cancer.
3. Key events: Measurable events that are critical to the induction of tumors.
4. Dose–response relationship: A discussion of whether the dose–response of the key events parallels the dose–response relationship of the tumor.
5. Temporal association: The key events should be observed before the tumor appearance.
6. Strength, consistency, and specificity of association of tumor response with key events: The weight of evidence linking the key events, precursor lesions, and the tumor response.
7. Biological plausibility and coherence: Consideration of whether the mode of action is consistent with what is known about carcinogenesis in general (biological plausibility) and in relation to what is known specifically for the substance (coherence).
8. Other modes of action: Discussion of alternative modes of action.
9. Assessment of postulated modes of action: Statement of the level of confidence in the postulated mode of action.
10. Uncertainties, inconsistencies, and data gaps: Uncertainties should include both those related to the biology of tumor development and those related to the database on the compound of interest. Inconsistencies should be flagged and data gaps be identified; gaps should be judged as to whether they are critical as support for the postulated MoA or just serve to increase confidence therein.

In order to provide guidance in determining the relevance of the MOA in animals for human risk assessment a human relevance framework concept (HRF) was developed by the International Life Sciences Institute/Risk Science Institute (ILSI/RSI) working group [25,26,134]. The HRF is based on the following four questions:

1. Is the weight of evidence sufficient to establish an MOA in animals?
2. Are key events in the animal MOA plausible in humans?
3. Taking into account kinetic and dynamic factors, is the animal MOA plausible in humans?
4. Conclusion: Statement of confidence, analysis, and implications.

The above process has been applied to several types of tumors and classes of compounds [25,86] and is a useful tool in conducting a human risk assessment based on a positive carcinogenicity study.

Examples of Rodent Tumors of Questionable Relevance to Humans

Mononuclear cell leukemia (MNCL) in F-344 rat: This tumor type is unique to the rat and is only common in the F-344 strain. King-Herbert and Thayer [113] reported a frequency in untreated F-344 rats in studies conducted

by the National Toxicology Program of 48.2% and 23.7% in males and females, respectively. Tumor development occurs only after an apparent threshold is exceeded. Some genotoxic carcinogens did not increase the incidence of MNCL, whereas several noncarcinogens did induce an increase. Therefore an increase in MNCL in F-344 rats is not considered relevant to humans [16].

α2μm-globulin associated renal tumors: These are male-rat-specific tumors occurring as a result of accumulation of a male-rat-specific protein, α2μm-globulin, in phagolysosomes of renal proximal tubular cells. As an analogous protein is not produced in humans those rodent tumors are not considered relevant for human risk assessment. The International Agency for Research on Cancer (IARC) has published a list of criteria that have to be met in order to support this mechanism of action [186].

Thyroid tumors rats: In rats, thyroid carcinogenesis can be induced by agents interfering with the pituitary-thyroid feedback mechanism. Hepatic enzyme inducers can increase thyroid hormone metabolism leading via a positive feedback mechanism to a stimulation of the thyroid gland. The same pathways also exist in humans, but there are some differences making the human more resistant to developing thyroid tumors. Species differences in thyroid physiology between rodent and human can explain the formation of thyroid tumors. Thyroxine-binding globulin (TBG) is the main human plasma protein that binds and transports thyroid hormone in the blood. Rodents are lacking this protein. In addition, the half-life of thyroxine is 16 h in rats versus 5–9 days in humans and serum levels are about 25 times higher in rodents than in humans, indicating higher activity of the rodent thyroid gland [194].

Urinary bladder tumors in mice and rats: Urinary bladder tumors can be induced through chronic irritation and subsequent increased cell proliferation followed by malignant transformation caused by precipitates. The same mechanism can also occur in humans if the chemical causing the formation of irritating objects is present in sufficient amounts. However, there are some physiological and anatomical differences between rodents and humans that make humans less susceptible. Rodent urine has high osmolality and a high concentration of protein compared to humans [24]. Calculi can more easily remain in the horizontal quadruped rodents compared to the upright walking humans [37]. Rodent bladder tumors are not relevant for humans if they only occur above a threshold concentration at which precipitation occurs.

Liver tumors in mice and rats: Many nongenotoxic chemicals produce liver tumors in rodents, especially in mice [81]. Proposed MOAs include cytotoxicity followed by persistent regenerative growth, enzyme induction, hormonal perturbation, immunosuppression, and porphyria [11,86,134]. A compound that causes liver tumors in mice only is frequently regarded as being of limited relevance to humans [17].

Hormonal disturbance: Disturbance of the hormonal balance is a common cause for induction of tumors in rodents, which is often due to the specific endocrine physiology of rodents and, therefore, without relevance to humans.

Leydic cell tumors: Various agents interfering with the hypothalamic–pituitary–testicular (HPT) axis and ultimately causing increased concentrations of serum LH (luteinizing hormone) have been shown to increase Leydic cell tumors especially in rats, but also in mice and beagle dogs (eg, androgen receptor agonists, 5α-reductase inhibitors, testosterone biosynthesis inhibitors, aromatase inhibitors dopamine agonists, gonadotropin-release hormone agonists, estrogen agonists/antagonists). The regulatory mechanisms of the HPT axis in rats and humans are similar, but humans seem to be less sensitive in their response to increased LH levels. Based on the fact that Leydig cell adenomas and carcinomas in the general population are very low and surveillance databases have detected no increased incidence it was concluded that human males are generally less sensitive than rodents. However, each situation has to be evaluated on a case-by-case basis [22,157].

Uterine tumors: Dopaminergic alkaloids have significant endocrine effects in rodents, particularly in rats, through their inhibitory effect on prolactin secretion [55]. Bromocriptin caused squamous cell metaplasia of the uterine endometrium in a chronic 53-week study in rats that progressed to uterine adenocarcinomas in the 2-year bioassay [56]. Normally older rats remain in diestrus with high prolactin and low LH levels. Lowering prolactin bromocriptine treatment initiated cyclic activity. However, a normal estrus cycle was not achieved and a higher estrogen/progesterone ratio led to the development of squamous endometrial metaplasia, which facilitated endometritis and pyometra and through irritation resulted in increased cell proliferation and finally to neoplasia. No such findings were detected in a 52-week dog study or in a carcinogenicity study in mice [56,153]. Endometrial biopsies of patients did not show any drug-related changes [164]. Therefore, the uterine changes in rats are without relevance for women and considered an exaggerated pharmacodynamic effect specific for aging female rats.

Target Organ Concordance Between Test Species and Human

Harderian gland (eye), Zymbal's gland (ear), preputial gland, clitoral gland, and forestomach are rodent specific organs that do not have a human equivalent, and hence, tumors in those organs are often regarded as not relevant to humans. However, target organ

concordance is not a prerequisite for the relevance of animal study results to human risk assessment. Physiological growth control mechanisms at the cellular level are similar among mammalian species, but these mechanisms are not necessarily site concordant. Specific considerations, however, may apply occasionally. For example, the rodent forestomach resembles the epithelium of the esophagus. Locally irritating substances may be rodent forestomach carcinogens through prolonged contact with the epithelium causing chronic irritation and inflammation. A carcinogenic risk for humans is considered unlikely, since exposure of the epithelium in the mouth, pharynx, and esophagus in patients swallowing a pill is short-lived. By contrast, the local exposure of the stomach of rodents treated by oral gavage is prolonged. Therefore exposure to nongenotoxic compounds at concentrations far below those having irritating potential is not a risk to human [208]. For other rodent-specific tumors a mode of action may not be easy to establish and a full weight of evidence approach has to be used in order to assess the risk to man, as described above [183].

The cases described above are only examples of tumors observed in animal studies that may not be relevant to humans. However, for new developmental compounds the hypothesis of a possible mode of action and the relevance to humans need to be supported by a weight of evidence approach for each case specifically.

Reproductive Toxicity Testing

Reproductive toxicity testing is a special area in preclinical safety because there are no dedicated follow-up studies in humans, ie, the aim of this part of the program is not to establish appropriate monitoring in humans but to identify potential hazards to reproduction based on which an integrated risk assessment is made to assess the possible impact of the observations in animal studies for humans. Based on this assessment, appropriate measures are to be implemented to manage and mitigate respective risks in humans with the ultimate aim to prevent adverse effects on all stages of human reproduction. These concepts are laid out in a number of guidelines across regions, some of which were issued recently including:

- ICH S5(R2). Detection of toxicity to reproduction for medicinal products and toxicity to male fertility [101].
- ICH M3(R2). Conduct of human clinical trials and marketing authorization for pharmaceuticals [90].
- FDA Guidance for industry. Reproductive and developmental toxicities—Integrating study results to assess concerns [69].
- EMA Guideline on risk assessment of medicinal products on human reproduction and lactation: From data to labeling [50].
- FDA Content and format for human prescription drug and biological products; requirements for pregnancy and lactation labeling [72].
- CTFG (Clinical Trial Facilitation Group) in Europe. Recommendations related to contraception and pregnancy testing in clinical trials [52].
- FDA Guidance for industry: Assessment of male-mediated developmental risk for pharmaceuticals (draft guidance) [73].

It is outside the scope of this chapter to explain the experimental methods used as to how the range of studies is being conducted, but detailed descriptions are laid out in the ICH S5-R2 guideline on the Detection of Toxicity to Reproduction for Medicinal Products and Toxicity to Male Fertility, 2005 and are discussed in chapter 9 of this book. The principal testing strategy should ensure exposure of all mature adults and all stages of development from conception to sexual maturity. To allow for detection of immediate and latent effects of exposure, observations should be continued through one complete lifecycle, ie, from conception in one generation through conception in the following generation. Exposures from weaning through puberty are not fully covered in the reproductive toxicity studies, and additional studies in juvenile animals should be considered, where appropriate; juvenile studies are also outside the scope of this chapter but are described in chapter 11 of this book. If several reproductive toxicity studies are conducted, it is mandatory to assure that no gaps in treatment occur, which can be determined by an overlap of at least one day in the exposure period of related studies.

The history of drug development shows how this field has developed in recent decades, and how we are still on a learning curve as to the predictivity of animal findings for humans in certain cases.

Thalidomide

Unfortunately, long before standard methods for the evaluation of developmental and reproductive toxicity (DART) were established and used routinely in the pharmaceutical industry as is the case today, and before the knowledge of experts in this field was well developed, thalidomide was discovered, received marketing approval in Germany in 1957, and was available over the counter from the beginning. The legal framework for the regulation of new medications was not comparable to our today's global standards, and these have also greatly developed since. Based on nerve damage in hands and feet, reported in elderly patients in 1959, Grünenthal applied step-wise for prescription-only status in selected German federal states in May 1961 [84]. In November 1961, the appearance of the very severe and typical malformations in

humans was associated with this highly potent human teratogen by two physicians, Dr. Widukind Lenz (in Germany) and Dr. William BcBride (Australia) and within 10 days, Grünenthal decided to withdraw thalidomide from the market [84].

Since then, reproductive toxicity testing became an integral part of preclinical safety testing. It is worth reviewing the thalidomide case to provide some insight as to how this tragedy has happened and to ask the question of how well we may or may not be protected today against a similar situation. James Schardein has described the history of thalidomide on the background of the general scientific understanding and attitude in the 1950–1960s in his chapter: "Thalidomide: The Prototype Teratogen" in *Chemically Induced Birth Defects*, 3rd edition, 2000 and highlights the fact that in these early days, reproductive toxicity testing was not an integral part of the safety assessment of medicines in development at all, and, moreover, that testing of fetal endpoints was generally missing from the testing paradigm. In addition, the knowledge about teratogenesis was not yet well developed then, and only 9 years before thalidomide had been identified as the cause of the human malformations in West Germany, it was known that drugs (ie, aminopterin) can cause human malformations when given during pregnancy. Dally [32] published an article in *The Lancet* about thalidomide to ask the question of whether the tragedy was preventable. This article highlights the fact that in spite of the available evidence demonstrating that fetal damage could occur through environmental influences, such as alcohol—already established in nineteenth century—this knowledge simply was largely forgotten by the mid-twentieth century. Medical students learned that the placenta was a barrier that protected the fetus up to doses that would kill the mother. Schardein highlights in his article that even leading teratologists at the time were skeptical about the association of thalidomide with human malformations [77,207].

Originally, thalidomide was studied for its anxiolytic, mild hypnotic, antiemetic, and adjuvant analgesic properties. Later it was found to be efficacious in the treatment of the cutaneous forms of leprosy (erythema nodosum leprosum) and since has been approved for the treatment of multiple myeloma; actually, it is considered to be potentially efficacious for the treatment of many more severe clinical conditions [49,155]. Its pharmacological mechanism of action is characterized by antiangiogenic and immunomodulatory properties. A number of mechanisms has been proposed for the former, including a downregulation of TNF-α, of vascular endothelial growth factor (VGEF) expression, the inhibition of response to basic fibroblastic growth factor (bFGF), and VEGF potentially through the modulation of integrin expression and impairment of migration, the inhibition of endothelial cell proliferation, and even blocking of cyclooxygenase-2 (COX-2). More mechanisms, however, are being evaluated, and thalidomide appears to influence many biological activities [49,109]. Many of these mechanisms play a fundamental role in physiology, but other drugs that interact with the same molecular pathways were not found to show comparable potential to cause a similar pattern of adverse effects on human embryo–fetal development.

The developmental toxicity of thalidomide in humans is characterized by typical congenital malformations, most prominently presenting as phocomelia, ie, stunted limbs, or the complete absence of limbs (amelia). Malformations may also affect the digits and hips or the ears, lips, palate, eyes, heart, spine, respiratory or gastrointestinal tract, and the urogenital system, ie, the kidneys or reproductive organs. Tragically, even a single dose of 50 mg was sufficient to cause the characteristic pattern of malformations, when taken during the critical phase of development of the limbs and the major organ systems, ie, during days 21–35 after conception [49,109,155]. This implies that a pregnant woman may not even have been aware of her pregnancy and/or could have been suffering from morning sickness against which the strong antiemetic properties of thalidomide were highly effective.

Since then, thalidomide has been extensively characterized in numerous species, strains and breeds ([175] and references cited within) including rats, mice, rabbits, dogs, hamsters, primates, cats, armadillos, guinea pigs, swine and ferrets. However, the pattern of adverse effects is greatly variable across species and a number of studies was negative, even for different strains of the same species. Thalidomide was found to be mostly embryo-toxic in the rat whereas rabbits and primates showed the best concordance with the typical human phocomelia. In addition, thalidomide is a much more potent teratogen in humans than in any of the animal species studied except the hamster and is much more toxic to the embryo than to the mother. Until today, the mechanism of developmental toxicity remains a mystery [175].

This example highlights the tragic combination of issues that together led to the most dramatic unexpected and adverse outcome in humans affecting such a high number of individuals:

- the misleading medical understanding of the nature of the placenta at the time that was considered to be an impermeable barrier to environmental influences;
- the testing paradigm of medicine in development at the time;
- the antiemetic efficacy of thalidomide in particular;

- the treatment of morning sickness, a common condition in pregnant women during the early and most vulnerable phase of gestation;
- the free availability over the counter due to much less restrictive legal regulations of new medicines entering the market;
- the fact that low and even single doses were sufficient to adversely affect embryo–fetal development;
- the much greater human sensitivity compared to most animals;
- the greater toxicity to the developing conceptus compared to the mother;
- and finally, the variable, often negative response in animals that hardly reflected the pattern in humans even after full knowledge of the human adverse effects.

It is of note that in rats, the prevailing outcome was characterized by embryotoxicity rather than teratogenicity, which highlights that the response in a biological system A may differ significantly from the response in a biological system B and yet still reflect a similar reaction to a common insult. Today, we do understand that *any* adverse effect on reproduction and development in animals may signify potential toxicity to these systems at unknown doses in humans, the development of which could have a very different phenotypic appearance. Therefore, a weight of evidence approach of all nonclinical safety studies is pursued to arrive at an integrated assessment of a potential reproductive or developmental risk for humans [50,69]. Based on the current understanding of thalidomide, a rabbit fertility and early embryonic development study was conducted by the applicant to support approval of thalidomide in multiple myeloma. This study demonstrated adverse effects on a number of parameters including an increase in resorptions. However, in prospective programs, such a study would not be conducted in rabbits but in rats instead and it is unclear whether the rat would have shown similar observations. It is not uncommon to see drugs in development with fairly unspecific, variable, and often mild outcomes in different species, as was the case for thalidomide and also for other compounds for which some more examples are discussed in the following.

For the new indications, treatment with thalidomide is highly regulated through a specifically developed risk evaluation and mitigation strategy (REMS), called Thalomid REMS [18]. The only purpose of an REMS managing a drug with a safety profile, such as of thalidomide is to prevent unintended exposure of a developing conceptus to a medication that is known or suspected to be developmentally toxic in humans. However, very strict measures have to be implemented and followed to achieve this goal and even under an effective REMS there remains a residual risk to expose pregnant women to thalidomide. Only in serious or even fatal conditions with little or no therapeutic alternatives may such risks be considered acceptable and the risk–benefit ratio still positive in spite of severe side effects, provided that effective measures can be implemented to minimize the associated risks.

Angiotensin-Converting Enzyme (ACE) Inhibitors

This class of compounds is indicated for the treatment of hypertension and congestive heart failure, with the first in class being captopril (Capoten approved in the United States in 1981) and the second in class being enalapril [203]. Enalapril may serve as an example here to illustrate the findings in humans and how they were missed in the toxicological studies.

Enalapril much in contrast to thalidomide was characterized in a full set of reproductive toxicity studies [203, 204], some of which used a modified design compared to the routine since the implementation of ICH in 1993 [101], particularly for the study for fertility and early embryonic development. That is, this study included a fetal evaluation and a lactation phase and was called "study of fertility and general reproductive performance." Since the implementation of ICH, in this type of study, usually, mated females are sacrificed around mid-pregnancy, which allows for evaluation of embryo toxicity but not teratogenicity or postnatal development.

In the rat embryo–fetal development study, maternal and fetal body weight development were reduced that could be prevented with the supplementation of pregnant dams with physiological saline, pointing to an underlying pharmacological effect. Enalapril was found to be neither teratogenic nor embryo-lethal. The rabbit study on embryo–fetal development showed no teratogenic effects either but maternal and embryo–fetal toxicity across the dose range tested, which could be prevented with the supplementation of physiological saline in the low-to-mid dose range but not at higher dose levels.

The rat study on peri- and postnatal development with treatment from day 15 of gestation to day 20 of lactation revealed reduced maternal and pup weight gain and an associated developmental delay for righting reflex, negative geotaxis, and landmarks of sexual developmental but no malformations. Behavioral assessments (open field and swimming maze) were unaffected. The reproductive phase showed no adverse effects on the F1 generation including their offspring, ie, litter size, the number of live and dead pups, or pup weight. The F2 pups revealed no external abnormalities. The ICH standard study design nowadays requires treatment from around implantation (ie, day 6 of gestation) to day 20 of lactation, which covers the major organogenesis in that study in addition.

In the rat study on fertility and general reproductive performance, males were treated from 70 days before mating, throughout mating and until termination of gestation of the corresponding female. Females were treated for 2 weeks prior to mating, throughout mating and gestation. One half was sacrificed on day 20 of gestation; their fetuses were subjected to skeletal (and assumed visceral) evaluation. Offspring allocated to this phase of the study showed reduced fetal weight. The other half of the dams was allowed to litter down and rear their offspring that underwent extensive postweaning examinations including sexual development and behavioral assessments; pup mortality was increased during the lactation phase and body weight gains in male pups were reduced after weaning. The F1 generation was allowed to mate and deliver; F2 litters were evaluated for litter size, numbers of live and dead pups, pup weight, and external abnormalities; these investigations revealed no effects on either the F1 of F2 generation. Fetal skeletal evaluations showed variations, ie, incomplete ossification of sternebrae and lumbar ribs. The description states that skeletal variations were not seen in F1 pups born normally, which implies a skeletal examination phase for delivered offspring, but in this respect, the details given are insufficient to confirm this aspect. In the F1 offspring, there were delays in the development of the surface righting reflex, auditory startle, and vaginal opening but no behavioral changes.

The overall pattern of observations therefore revealed unspecific and fairly mild observations that are not uncommon in this type of study, but with confounding maternal toxicity. Saline supplementation was preventive, which seems to point to a pharmacologically mediated effect. The signal was most evident in the F1 generation from the study on fertility and general reproductive performance. Overall, the combination of findings could be interpreted to indicate a pattern of "developmental delay secondary to maternal toxicity and/or pharmacological effects" that was not deemed too concerning. There was no indication of a primary dysmorphogenic mode of action; the findings are more likely secondary effects mediated through the primary mode of action on the dams and/or the offspring.

Unfortunately, human data demonstrated unexpected and serious concerns, particularly becoming evident during the second and third trimester as intrauterine growth retardation and an increased risk of fetopathy, presenting as renal dysplasia, renal failure, anuria and death, oligohydramnios, and specific adverse outcomes secondary to amniotic fluid volume, ie, limb deformities, cranial ossification deficits, and lung hypoplasia. In addition, neonatal renal failure was observed. Fetal urine production in humans starts toward the end of the first trimester [13,28,188].

This adverse human outcome was totally unexpected from the comprehensive preclinical studies, and it is worth reviewing the designs that were used. Human risk was most evident when treating women with hypertension—a common complication in pregnancy—during the second and third trimester of pregnancy, but not during the first. In the study on peri- and postnatal development, treatment started on day 15 of pregnancy, which is slightly before the end of organogenesis and best reflects the clinical treatment conditions leading to adverse human outcomes. However, there was no indication for such a severe adverse effect in this study. The question therefore is how the difference can be explained and why the animal studies failed to predict the risk in humans.

Tabacova and Kimmel [187] reviewed the typical ACE-inhibitor induced adverse fetal outcome termed ACEI fetopathy. In humans, the target system of enalapril, ie, the kidney and the renin–angiotensin system, develops at the end of the first trimester and prior to skeletal ossification. In most of the animal species studied, the enalapril target systems are comparably less mature, and consequently, enalapril cannot work on them until they are functional. Only shortly before term are these systems developed, at which point the fetus is more mature and less vulnerable. In particular, the rat shows greatest disparity with respect to the relative development of the kidney and skeletal ossification compared to humans, which explains why effects similar to humans were not detected in the rat reproductive toxicity program, in spite of some apparent pharmacological effects. Rhesus monkeys show the best concordance of the prenatal development of these systems with humans, but this species is not routinely used in embryo–fetal development studies unless there is a specific justification. For the testing of ACE inhibitors, indeed, the use of the rhesus may have been the better choice, but this was not evident at the time of prospective testing of this new class of compounds. Tabacova also pointed out that it is unclear whether a similar pathology would be seen in this animal model. It appears that exposure to enalapril after the first trimester was strongly associated with oligohydramnios and the specific adverse outcomes were considered secondary to the reduced amniotic fluid volume, as well as with neonatal renal failure [188]. Tabacova [189] concluded that "animal studies that follow standard protocols and evaluate developmental toxicity only for exposures during embryogenesis will miss developmental effects arising secondary to disruption of target systems that develop after the period of major organogenesis. Thus, although the animal mode of action (MOA) for enalapril and other ACEI is plausible in humans, differences in the timing of development of critical target organ systems, particularly the renal system and renin–angiotensin system (RAS), explain the absence of definitive

structural abnormalities in test animals." This example highlights again how the absence of evidence is not evidence of absence, and that a profound understanding of the test system is key for the interpretation of results.

Since the discovery of the adverse effects brought about by the ACE inhibitors when treating pregnant women during the second and third trimester of pregnancy, it also became a matter of debate whether ACE inhibitors could be potential teratogens when given during the first trimester. Angiotensin II receptors are widely expressed in fetal tissue [4] and Cooper et al. [28] suggested an increased risk of major congenital malformations, particularly of the heart and CNS. This hypothesis has stimulated a dialogue, but the answer to this question remains unresolved at present due to conflicting evidence [123,206]. There are confounding factors that complicate the assessment in a clinical setting, including obesity, diabetes, the hypertension itself, or other antihypertensive medications [123,159,173]. Sealey and Itskovitz-Eldor [176] commented that it is unknown whether the postulated effects are specific to ACEI or could be applicable to other drugs that block the RAS (eg, beta-blockers, ACE receptor blockers, renin inhibitors) [120], given that the oocyte, embryo, and developing fetus are continuously "bathed" in "prorenin, the precursor of renin, from just before ovulation until parturition" [80,106,107]. This aspect is particularly interesting in view of the more pronounced but still unspecific findings in the study on fertility and general reproductive performance as opposed to the peri- and postnatal study and in view of the complete absence of evidence for teratogenicity in the reproductive toxicity studies, which is not readily explicable if prorenin is a key determinant in embryo–fetal development. Since 2012, enalapril and other drugs, such as aliskiren [191], which fall into this category, have received a boxed warning in the United States indicating that "Drugs that act directly on the renin-angiotensin system can cause injury and death to the developing fetus." This statement is, notably, based on a hypothesis that yet needs to be confirmed, but is the current basis to aid in human risk management.

Endothelin Receptor Antagonists

The endothelin receptor antagonists were discovered in the late 1980s, with the first in class being bosentan (Tracleer), a mixed antagonist of endothelin receptors (ET_A and ET_B), which entered clinical development in 1993 and was approved as orphan drug for the treatment of pulmonary arterial hypertension in 2001 [23,195]. Generally, from its pharmacology, bosentan was believed to be a promising candidate for the management of clinical disorders associated with vasoconstriction. Among these, migraine was also evaluated clinically but bosentan was reported not to be effective [129].

In the course of development, bosentan was found to be teratogenic and embryo-toxic in the rat when given orally at doses as low as two times the MRHD based on body surface area [198,199]. The findings encountered included craniofacial abnormalities, such as agenesis of the palate, shortened/misshapen mandibles, fusion of the pterygoid process with the tympanic annulus, abnormal zygomatic arch, shortened tongues, anophthalmia, and microphthalmia. Blood-vessel findings were also observed. Similar observations could be demonstrated in a mouse knockout model, and these effects were more pronounced when pregnant dams were treated with other agents antagonizing endothelin or the ET_A receptor in addition [118,119]. Regulatory studies with bosentan in the rabbit, however, failed to show evidence of teratogenicity [198]. The only findings observed in this species were an impaired fetal body weight in the presence of maternal toxicity only and a higher incidence of some skeletal variations in the high-dose group. Hence, the rabbit is less sensitive than the rat in this case, and testing in the rabbit only would have resulted in a false-negative outcome, although it was noted that systemic exposures in the rabbit were lower than in the rat. It is of note that the high dose was 1500 mg/kg/day, which exceeds the limit dose in the ICH S5 guideline (see Table 2.1) [198,199]. The difference in exposures may explain the variable response between the animal species, since the pattern of findings in rats and knock-out mice strongly suggests a class effect associated with the mode of action. Indeed, other endothelin antagonists in development were published to cause a concordant pattern of malformations in rats and rabbits and the authors concluded that teratogenicity is a likely class effect of endothelin receptor antagonists [200].

In general, variable outcomes between species may be a matter of specificity or of different levels of sensitivity—also in a broader sense, ie, encompassing not only lower species sensitivity but also reduced sensitivity of the testing conditions—particularly in cases like this one where it would be difficult to understand why such a fundamental physiological target, which is involved in embryo–fetal development, does not result in a similar phenotypic outcome when inhibited. Again, macitentan (Opsumit), another endothelin antagonist approved for pulmonary arterial hypertension in the United States in 2013 [149,150], was demonstrated to show similar effects as bosentan in the rat, and was also teratogenic in the rabbit at all doses tested with both species showing a similar pattern, evidenced as cardiovascular and mandibular arch fusion abnormalities. Administration of macitentan to female rats from late pregnancy through lactation caused reduced pup survival and impairment of male fertility of the offspring at all dose levels tested [150]. Therefore lactating women should either

discontinue macitentan or nursing. The same precaution is also recommended for bosentan [199].

The example of the endothelin antagonists demonstrates the importance and obligation to manage potential human risks in the most responsible manner, ie, it must be postulated that the adverse effects established in the DART studies are predictive for humans, even in cases where only one species seems to show adverse effects. Potential risks need to be mitigated accordingly. Therefore in this case, again, unintended exposure of pregnant women must be avoided. In practical terms, the Tracleer [bosentan] label contains a boxed warning indicating the following: "Based on animal data, Tracleer is likely to cause major birth defects if used during pregnancy." Tracleer is only available through a restricted distribution program called the Tracleer Access Program (T.A.P.) because of this risk (and the risk of liver failure). Macitentan is handled accordingly. The applied risk management strategy is—notably—based on clear evidence of developmental risk to the unborn from regulatory animal studies and—thankfully—not on clinical evidence from epidemiological data, and this shows how the sound nonclinical characterization allowed for a meaningful risk–benefit evaluation based on which the risk can be managed effectively. However, it also is evident that bosentan or other endothelin antagonists are not approvable for "soft(er)" indications where alternative medications are available and/or where the risk of exposure of pregnant women is much greater, such as migraine, and therefore the risks outweigh potential benefits. By contrast, pulmonary arterial hypertension is a fatal condition and a very rare disease for which bosentan was originally granted an orphan designation. The potential benefit for this condition therefore was deemed to outweigh potential risks, when successfully managed. To the best of our knowledge, indeed, clinically, no case of congenital malformations associated with the use of endothelin antagonists has been reported to date.

Triptanes

In 1995, the first triptane, sumatriptan, received approval in the United States. This new class of compounds targeting serotonin receptors was developed for the treatment of migraine.

This indication affects all populations, including children, peaks around the age of 40, and then declines (data from the US). Females are more frequently affected than males. The age and gender distribution demonstrate that in this indication women of childbearing potential represent a great proportion of the target patient population, which makes reproductive toxicity assessment a key determinant in the safety assessment of new drugs in this field [125].

Sumatriptan, the first triptane in class, may serve as an example to tell the history of success of this innovative class of medicines with respect to the characterization of its developmental toxicity profile.

Sumatriptan was evaluated in a full range of regulatory studies using different routes of exposure, including intravenous (i.v.) and oral and was found to be embryolethal in rabbits when given daily i.v. at doses approximating the maximum recommended single human subcutaneous dose of 6 mg on a body surface area basis (MRHD). The doses were at or close to those producing maternal toxicity. Fetuses of pregnant rabbits administered oral sumatriptan (at doses greater than 50 times the MRHD) during organogenesis had an increased incidence of cervicothoracic vascular and skeletal anomalies. In contrast, embryo–fetal lethality was not observed in pregnant rats treated throughout organogenesis with i.v. doses approximately 20 times the MRHD. Moreover, no rat embryo–fetal lethality or teratogenicity was observed with daily subcutaneous doses before and throughout gestation [105]. Shepard [178] described a study in which no fetal adverse effects were observed in rats given up to 1000 mg/kg orally during organogenesis.

Following approval, the Sumatriptan/Naratriptan/Treximet Pregnancy Registry was established to monitor pregnancy outcomes following treatment with these medications. The interim report summarizing data from the 1st of January 1996 through the 31st of October 2011 was issued in May 2012 and concluded for sumatriptan that the "data do no indicate a signal for major teratogenicity." Data on Naratriptan and Treximet were too limited for the registry to meet its primary objective.

This example highlights several important aspects in the prospective risk assessment of new drugs—particularly for new classes of compounds. From the preclinical data set, clearly there was concern as to whether the data might signify an adverse effect of sumatriptan on embryo–fetal development in a particularly vulnerable target patient population. The findings were—as for thalidomide (rat), the ACE inhibitors, and many drugs in development (personal experience)—fairly unspecific and did not demonstrate a defined pattern as for the endothelin antagonists. In addition, maternal toxicity was a confounding factor. Obviously, however, there is a clear need to distinguish the hazardous compounds from those that are benign. In this case, a new class of compounds with major benefit for the patients affected was approved but at the same time, any potential risks were prospectively, carefully, and successfully managed, which is an important element in the development of innovative and beneficial drugs.

Overall, the field of reproductive toxicology has developed into a mature discipline since the occurrence of the thalidomide tragedy. Schardein et al. [174] have reviewed species sensitivities and the prediction of teratogenic

potential based on the observation that many xenobiotics shown to be teratogenic in animals are not known to be teratogenic in humans. The reasons for this apparent difference remain to be established and could involve a number of factors, including, indeed, lower species sensitivity of humans, but also subteratogenic exposure levels or a lack of appropriate methods to identify human teratogens. By contrast, with the exception of the coumarin anticoagulant drugs, all well-accepted human teratogens were also demonstrated to be teratogenic in at least one laboratory animal species. However, there is no single species that in general is giving a more reliable response than another; in fact, rats and mice showed the best concordance for findings observed in humans but in other cases also produced the most nonconcordant responses, whereas rabbits were more unlikely to give a false-positive response. Primates in general are showing higher predictivity levels but are less commonly used. In essence, the authors concluded that neither a single species nor a single study will be sufficient to detect a potential reproductive hazard but all endpoints must be taken into account, including results from other toxicology studies, pharmacokinetic and metabolic data as well as the pharmacological mode of action to arrive at an integrated assessment of human risk.

Maternal toxicity as a confounding factor is a matter of ongoing debate in the scientific community. Khera [112] has proposed this concept to put adverse developmental findings occurring at maternally toxic dose levels into context. While maternal toxicity is an important consideration and may well be a contributory factor to the development of unspecific effects in the offspring—such as a reduction of fetal weight at dose levels that simultaneously significantly impair maternal body weight development—specific patterns of malformations are highly unlikely to be secondary to maternal toxicity [9,33]. On the other hand, typical but rare malformations may be observed in a given strain due to genetic liability. In such a case, the actual incidence in a given study as compared to background data and aspects, such as the distribution across groups and a potential association with dose levels—for example, if malformations typical for this strain become evident only at a maternally toxic doses—are critical aspects and, depending on the outcome, may or may not increase concern with regards to potential developmental toxicity of a given test item. It is therefore critical to have a robust set of background data to be able to put findings in context on a case-by-case basis. The more limited a set of background data and the higher the incidence of background observations the lower is the sensitivity of a test system in a given laboratory. As a basic principle, it is important to assess whether adverse effects on the conceptus—be they specific or not—are observed in the presence of maternal toxicity only or whether the conceptus appears to be more vulnerable than the dam, which is an especially hazardous situation. Obviously, this correlation may not be the same in another species.

The current nonclinical tools to assess a developmental and reproductive hazard can be considered to be fairly effective but, inevitably, an intrinsic residual risk of failure remains given the biological complexity of reproduction, involving the closely intertwined maternal and embryo–fetal systems and their manifold interactions, which also show species-specific features. For the purpose of human risk assessment, the most challenging situations are those where the experimental data are inconclusive, eg, due to unspecific findings associated with confounding factors, such as maternal toxicity, a lack of concordance between species, or a lack of biological plausibility. Human evidence may either increase concern, such as in the case of the ACE inhibitors, or, alternatively, decrease concern, such as in the case of the triptanes. In some cases, projects with such inconclusive profiles will be terminated and not developed further.

It is virtually impossible to definitively confirm the absence of an adverse potential on embryo–fetal development in humans. Approximately 3% of newborns have congenital malformations requiring medical intervention, with about one-third being life-threatening. More than twice as many are detected later in life [179]. This demonstrates that any human teratogen would likely have to occur at a distinctly higher incidence, perhaps in a cluster or to be of a very unusual type, to be identified. History shows that this was the case indeed, eg, for the detection of thalidomide or diethylstilbestrol (DES) as human teratogens [174].

With regards to human safety, specific patterns of adverse developmental effects in animal studies must be considered to be predictive a priori and potential risks be managed appropriately to prevent harmful human outcomes, such as with the REMS in place for thalidomide or bosentan. In such cases, the severity of the human condition and the unmet medical need for treatment will determine whether potential benefits still outweigh potential risks in the target patient population.

POSTMARKETING

After years of comprehensive nonclinical and clinical characterization, finally, a new therapeutic made it all the way and got marketing approval. The launch was successful and large patient populations treated. The development costs can be recuperated und money can be made to support new research and development. However, in spite of well-conducted nonclinical and clinical studies, the unexpected can still occur at this stage and may even result in the withdrawal of prescription drugs from the market. Such situations are not rare events.