AAPS Advances in the Pharmaceutical Sciences Series 5

Jan Willem van der Laan
Joseph J. DeGeorge *Editors*

Global Approach in Safety Testing

ICH Guidelines Explained

aapspress
Springer

AAPS Advances in the Pharmaceutical Sciences Series

The AAPS Advances in the Pharmaceutical Sciences Series, published in partnership with the American Association of Pharmaceutical Scientists, is designed to deliver well written volumes authored by opinion leaders and authoritarians from around the globe, addressing innovations in drug research and development, and best practice for scientists and industry professionals in the pharma and biotech industries. For more details and to see a list of titles in the Series please visit http://www.springer.com/series/8825

Jan Willem van der Laan • Joseph J. DeGeorge
Editors

Global Approach in Safety Testing

ICH Guidelines Explained

aapspress Springer

Editors
Jan Willem van der Laan
Section on Pharmacology, Toxicology
and Biotechnology
Medicines Evaluation Board
Utrecht, The Netherlands

Joseph J. DeGeorge
Drug Safety Testing
Merck Research Laboratories
West Point, PA, USA

The opinions stated in this book are solely those of the authors and do not reflect the opinions of their employers or of the ICH.

ISSN 2210-7371 ISSN 2210-738X (electronic)
ISBN 978-1-4614-5949-1 ISBN 978-1-4614-5950-7 (eBook)
DOI 10.1007/978-1-4614-5950-7
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2012956298

Picture credit: zhigong 志功 Zhang/fotolia.com

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

As an initiator and cofounder of the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human use, I am pleased to see that this famous "ICH" initiative is still alive and well. I had the privilege of participating directly in this unique process until 2000, as a representative of the European Commission and later as the first director of the European Medicines Agency (EMA). I also participated in the launch of a similar initiative for the harmonization of testing requirement for veterinary medicines (VICH).

During the 1990s, most of the basic ICH guidelines were finalised, thanks to massive organisational efforts, good technical preparation and sound scientific debates between experts from the USA, Europe and Japan. Since 2000, the ICH work has shifted to maintenance, dissemination and formatting of the technical registration dossier.

I would like to pay tribute to the scientific excellence, teamwork and dedication demonstrated during all these years by the best international pharmaceutical experts. The background, motivations, agreements and disagreements around ICH topics were openly discussed during six major ICH conferences held between 1991 and 2003. Complete official proceedings of the first four conferences were published[1] and widely disseminated to experts worldwide.

In my view, recent ICH activities deserve better publicity. I was therefore very interested to learn that Jan Willem van der Laan and Joseph J. DeGeorge had been able to bring together the reflections of the best ICH safety experts in order to explain and illustrate the ICH approach in safety testing and beyond. This book provides a useful insight into the detailed discussions, and sometimes heated debates, amongst experts, as in all scientific advances. I am pleased to preface these important contributions with a few words about the origins, expectations and achievements of ICH.

[1] IFPMA, edited by D'Arcy and Harron, The Queen's University of Belfast, Greystone Books, Antrim, Northern Ireland, 1992, 1994, 1996, 1998.

From European to International Harmonization

During the 1980s, as the head of a small dedicated team at the European Commission in Brussels, I was in charge of European pharmaceutical harmonization, including drug testing requirements. The technical aspects of binding legislation (European Directives) and detailed guidelines were drafted with the help of an advisory scientific committee called the Committee for Proprietary Medicinal Products (CPMP) and its working groups on quality, safety, biotechnology and efficacy. All our experts were nominated by the Member States of the European Union and spent most of their time in national drug agencies, public laboratories or university hospitals.

Since most of pharmaceutical research and development is performed by, or on behalf of, the industry, I organised a regular forum for exchanges between experts nominated by the European Industry Federation (EFPIA) led by Nelly Baudrihaye and our CPMP experts. In parallel, we launched systematic six monthly public consultations on draft guidelines with health professionals, consumers/patients and scientific societies in order to attract a maximum number of comments, which were analysed and discussed within the CPMP structures.

I accompanied various European delegations involved in trade negotiations, especially with the USA and Japan. Most of the time, these discussions were of a general and diplomatic nature, with few tangible results. Therefore, I took the initiative to supplement these diplomatic missions with more technical discussions with the Japanese Ministry of Health or the US Food and Drug Administration. I asked the chairs of the CPMP working parties to accompany me: Pr. Jean-Michel Alexandre (efficacy), Dr. John Griffin and Pr. Rolf Bass (safety), Pr. Giuseppe Vicari and Manfred Haase (biotechnology) and Tony Cartwright and Jean Louis Robert (quality).

I took advantage of the International Conference of Drug Regulatory Authorities (ICDRA) organised under the auspices of the World Health Organization (WHO) to consult various influential regulators on what could be achieved multilaterally rather than on a time-consuming bilateral basis. The fourth ICDRA in Tokyo, 1986, was such an occasion. I tried to convince the Japanese side and the FDA of the merits of a trilateral initiative. A formal green light was given to me by the US and Japanese colleagues in the margins of the fifth ICDRA, Paris 1989.

Starting the ICH Process

I had the honour to chair the very first ICH steering committee meeting in Brussels in April 1990. We discussed in a friendly atmosphere our general and specific objectives and the necessary consultation process, including public conferences. It was felt that European harmonization provided a good model. We agreed to invite WHO in order to facilitate the consultation and involvement of the rest of the world. In 1992, the 45th World Health Assembly adopted a resolution supporting the ICH initiative. We also invited Canada and the European Free Trade Association to send an observer to all our meetings.

Together with Osamu Doi, Tatsuo Kurokawa, Isamu Shimada, Elaine Esber, Alex Giaquinto, Nelly Baudrihaye and Richard Arnold, we identified four major objectives for ICH:

- To establish constructive scientific dialogue on the differences in registration requirements
- To identify areas of mutual acceptance of research results without compromising safety
- To recommend practical ways to achieve greater harmonization of registration requirements
- To reduce the unethical repetition of tests in animals and humans

We agreed to involve the research-based industry on a consultative basis, making clear that regulators were in control of the agenda, had a decisive role at every step of the ICH process and were solely responsible for finalisation. We also wanted to discuss the scientific issues in large public meetings, involving the best scientists from all over the world. The three regional research-based industry representatives suggested that their international federation (IFPMA) would take care of secretariat, logistics and publication, and Margaret Cone was put in charge. Nelly Baudrihaye agreed to take the huge risk of inviting more than 1,000 experts to the first public conference (plenary and breakout sessions).

Before the first major public event (ICH1 Brussels, Nov. 1991), we held 3 additional steering committee meetings, flanked by expert working groups in order to agree on the ICH goals, start work on the first 11 topics (quality, safety, efficacy and multidisciplinary aspects), refine the procedural steps and prepare the future ICH programme.

Without signing any formal international agreement, the three regulators pledged to bring present the ICH results to their respective national or regional authorities, in conformity with their local procedures.

Any difficulties would be monitored and discussed at subsequent public event. In my view, transparency and good regulatory governance were essential prerequisites to secure the success of our common enterprise. To this end, I had to create a dedicated Commission fund to cover the travel expenses of national regulatory experts from the CPMP and its working parties. Travel costs of speakers at major ICH public conferences were covered by attendance fees.

Impressive Results for ICH, So Far

I had an opportunity to express my personal views during the plenary sessions of ICH1 (Brussels), ICH2 (Orlando), ICH3 (Yokohama) and ICH5 (San Diego). I also attempted to summarise the early achievements of the ICH process in a collective booklet[2] on ICH published under the direction of Professor Jose Luis Valverde, member of the European Parliament.

[2] "The story of ICH", European Pharmaceutical Law Notebooks, Vol II, N° 4, May 1996, CEFI/CEDEF, Madrid

The ICH process, which has now been in operation for more than two decades, has proved to be an ongoing success. Thanks to the practical and concrete approach from all interested parties, more than 50 trilateral guidelines have been adopted and published on the Web (http://www.ich.org). Most ICH goals have been achieved, and the focus is now on maintenance and dissemination activities. Since 2006, several safety guidelines have been revised to incorporate new scientific knowledge.

Questions have been raised regarding the legitimacy and accountability of the ICH process. The fact is that ICH does not have the status of decision-making international institution. It remains a public/private advisory forum, seeking scientific consensus. Therefore, ICH has no authority to impose regulatory requirements on the three regions. It can only advise regulators and regulated on scientific issues. Each region must consult their relevant authorities and interested parties according to their own sovereign procedures. Each regulator can therefore block the ICH process at any stage.

In Europe, for example, ICH draft guidelines have to be circulated for European-wide consultations. At their final stage, they are scrutinised and adopted by the EMA's scientific committees. The ICH parties redefined the principles of governance in June 2012, confirming, clarifying and reinforcing the role of regulators in the process.

The difficulties in finding a common approach on certain topics such as the extent of repeat dose toxicity needed to start clinical trials or the duration and number of animal species for carcinogenicity studies were openly discussed during successive ICH public conferences. The issues at stake and the solutions found are fully addressed in this book.

With the support of ICH and its Global Cooperation Group, several regional harmonization initiatives have taken place, as well as training and feedback. I witnessed the enthusiasm and great potential of such initiatives when I chaired the third African Regulatory Conference organised in Accra (Ghana) by the Drug Information Association in May 2012. Several regional organisations in Africa are trying to bring together their limited regulatory resources, taking inspiration and advice from the US FDA and the European Medicines Agency and making reference to ICH guidelines.

This book provides a general perspective from the three main regulators as well as an overview of current scientific thinking and trends on preclinical issues such as toxicokinetics, duration of toxicity testing, carcinogenicity, reproduction and genotoxicity testing, safety pharmacology and safety evaluation of biotech products. It also covers the concrete efforts made to reduce animal experiments without compromising the safe development of new treatments.

The authors have succeeded in showing how in-depth understanding between Europe, the USA and Japan has been built up. In stimulating comparative evaluation of datasets, ICH has contributed to a better definition of what is needed in terms of safety testing. The areas still open for future debates are also described. This book leaves the reader with a distinct feeling of optimism about the future of good regulatory practices and international cooperation between regulatory authorities, for the benefit of patients worldwide.

France Fernand Sauer

Contents

Contributors

Rolf Bass Pharmaceutical Medicine, University of Basel, Basel, Switzerland

Pharmacology and Toxicology, Charité, Berlin, Germany

BfArM, Bonn, Germany

Boelckestrasse 80, Berlin–Tempelhof, Germany

Bruce Campbell Department of Pharmacology, Kings College Guys Campus London, Proximagen Neuroscience, London, UK

Joy Cavagnaro Access BIO, Boyce, VA, USA

Joseph J. DeGeorge Drug Safety Testing, Merck Research Laboratories, West Point, PA, USA

Sheila Galloway Merck Research Laboratories, West Point, PA, USA

Kenneth L. Hastings Sanofi SA, Bethesda, MD, USA

Makoto Hayashi Biosafety Research Center, Foods, Drugs and Pesticides (BSRC), Shioshinden, Iwata, Shizuoka, Japan

Bob Ings RMI-Pharmacokinetics, Carlsbad, CA, USA

David R. Jones Medicines and Healthcare products Regulatory Agency (MHRA), Victoria, London, UK

John E. Koerner United States Food and Drug Administration, Center for Drug Evaluation and Research, Silver Spring, MD, USA

Jan Willem van der Laan Section on Pharmacology, Toxicology and Biotechnology, Medicines Evaluation Board, Utrecht, The Netherlands

John K. Leighton Office of Oncology Drug Products, Center for Drug Evaluation and Research, US Food and Drug Administration, Silver Spring, MD, USA

Jonathan Moggs Discovery & Investigative Safety, Preclinical Safety, Novartis Institutes for Biomedical Research, Basel, Switzerland

Justina A. Molzon Center for Drug Evaluation and Research, U.S. Food and Drug Administration, Silver Spring, MD, USA

Lutz Müller Non-Clinical Safety, F. Hoffmann-La Roche Ltd., Basel, Switzerland

Yasuo Ohno National Institute of Health Sciences, Setagaya, Tokyo, Japan

Klaus Olejniczak Scientific Director in the Federal Institute for Drugs and Medical Devices (BfArM), Head Geno-and Reproductive Toxicity Unit (retired), Berlin, Germany

Hiroshi Onodera Senior Scientist (Toxicology) Pharmaceuticals & Medical Devices Agency, Tokyo, Japan

Fernand Sauer Former Executive Director of the European Medicines Agency, Cassis, France

Peter K.S. Siegl Siegl Pharma Consulting LLC, Blue Bell, PA, USA

Jennifer Sims Integrated Biologix GmbH, Basel, Switzerland

Frank Sistare Drug Safety Testing, Merck Research Laboratories, West Point, PA, USA

Per Sjöberg Eureda KB, Uppsala Science Park, Uppsala, Sweden

Per Spindler Biopeople, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

David Tweats The Genetics Department, The School of Medicine, University of Swansea, Swansea, UK

Beate Ulbrich BfArM, Bonn, Germany

BFR, Berlin, Germany

Spiros Vamvakas European Medicines Agency, London, UK

Herman Van Cauteren Pharmaparacelsus LLC, Gierle, Belgium

Chapter 1
The International Conference on Harmonisation: History of Safety Guidelines

Jan Willem van der Laan and Joseph J. DeGeorge

Abstract The International Conference on Harmonisation started in 1989. An overview has been given about milestones and history of the progress, focused on the safety topics.

Some reflections have been given about differences in approaches between the authorities involved and the legal bases that in some cases have also led to differences in scientific approaches. Harmonization of guidance content is much easier than harmonization of the systems.

Looking back on the ICH process after more than 20 years can teach us that by investing significant effort, an in-depth mutual understanding has been reached between the various regions and also between pharmaceutical industry and regulators.

1.1 Start of ICH

The history of ICH started in 1988 when a delegation of the European Commission with European Pharmaceutical Industry visited Japan. During this visit, differences in technical requirements for pharmaceutical for human use were identified as being a stumbling block for further cooperation between these two economic regions in the world.

These differences between regulatory agencies were challenged because the agencies in their own region had the same responsibility, i.e., ensuring the safety, quality, and efficacy of the medicines for humans on their respective markets.

J.W. van der Laan (✉)
Section on Pharmacology, Toxicology and Biotechnology, Medicines Evaluation Board, PO Box 8275, 3503RG, Utrecht, The Netherlands

Graadt van Roggenweg 500, 3531AH, Utrecht, The Netherlands
e-mail: jw.vd.laan@cbg-meb.nl

J.J. DeGeorge
Drug Safety Testing, Merck Research Laboratories, West Point, PA, USA

J.W. van der Laan and J.J. DeGeorge (eds.), *Global Approach in Safety Testing*, AAPS Advances in the Pharmaceutical Sciences Series 5, DOI 10.1007/978-1-4614-5950-7_1,

Table 1.1 Sixpack: International conference on harmonisation

Regulatory parties	Industrial parties
1. European Union (including European Medicines Agency and CHMP)	4. European Federation of Pharmaceutical Industry Associations (EFPIA)
2. US Food and Drug Administration	5. Pharmaceutical Research and Manufacturers Association (PhRMA)
3. Japanese Ministry of Health, Labor, and Welfare (now including PMDA)	6. Japanese Pharmaceutical Manufacturers Association (JPMA)

Secretariat: International Federation of Pharmaceutical Manufacturers Association

The concept was raised by the industry that by reducing these differences the costs of developing promising new pharmaceuticals could be reduced.

The project was elaborated further not only between Japan and the European Community but also with the USA and its regulatory authority Food and Drug Administration (FDA) with its Center for Drug Evaluation and Research (CDER) and its Center for Biologics Evaluation and Research (CBER). In October 1989 in Paris, the project received green light to proceed.

Small detail: This is the reason that the word "Harmonisation" is spelt with an "S" in the British way, highlighting the European starting point. The full name is International Conference of Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use.

The participants starting the ICH process are listed in Table 1.1.

Observers are invited (and were present) from other regulatory authorities (European Free Trade Association [EFTA], e.g., Sweden representing also Switzerland, and Norway; Health Canada; Therapeutic Goods Administration [Australia]; and World Health Organization [WHO]).

The European Community at that time consisted of 12 member states. Sweden, Finland, and Denmark joined in 1995, just preceding the establishment of the European Agency for Evaluation of Medicines (EMEA) per second half of 1995.

A steering group was established and a first meeting being held in Brussels in April 1990. Three other meetings took place in Tokyo (1) and Washington (2), and parallel meetings of working groups were organized to work on the various technical guidance documents.

In October 1990, the initial steering committee meeting in Tokyo published a statement expressing their commitment to increased international harmonization to ensure that good quality, safe, and effective medicines are developed and registered in the most efficient and cost-effective manner.

In 1991, the First International Conference on Harmonisation (ICH) was held in Brussels with more than 1,000 participants. The full list of the biannual (later triannual) meetings is given in Table 1.2.

These meetings were held in conjunction with working group meetings and were used to broadcast the accomplishments derived from the ICH activities occurring since the last general forum. This formal roll-out ceased after the Osaka meeting,

Table 1.2 International conferences on harmonisation being held

	Year	Place
ICH 1	1991	Brussels
ICH 2	1993	Orlando
ICH 3	1995	Yokohama
ICH 4	1997	Brussels
ICH 5	2000	San Diego
ICH 6	2003	Osaka

but the working group and steering committee meetings and accomplishments continue. The presentations of first four ICH meetings were recorded precisely and published in books with a steady increase in the number of pages doubling over the four conferences (D'Arcy and Harron 1992, 1994, 1996, 1998). ICH 5 was reported by the Regulatory Affairs Journal with Margaret Cone as the editor (Cone 2001). Mrs. Cone was during the earlier ICHs leading the IFPMA secretariat and knew the ICH process from inside.

The sixth ICH held in Osaka was no longer reported in a book.

The ICH 5 and ICH 6 reports are available on a CD in line with the technical developments in that period.

1.2 Organization of ICH

The ICH is not an authority. After agreeing on a potential expert topic needing harmonization between the six parties, a concept paper is to be written, and the steering committee (SC) has to agree with this goal, its potential impact on drug development, and the timelines for achieving the guidance development steps. Once the SC agrees, an official Expert Working Group (EWG) can be formed with representatives from all parties. One of the parties will act as the rapporteur, in most cases the initiator of the activity and, usually at the early stage of the project, a representative from industry.

The working process is a 5-step process (Fig. 1.1) starting with the first initial agreement, identifying the guidance and harmonization gaps (step 1) and coming to a first full agreement that is to be published (step 2) as a paper to be signed off by all six parties. Public consultation is step 3, and this is organized by the individual regulatory parties via their common way of communicating with the outside world. Comments are to be received via this route. In step 3, the comments are gathered and sent to the EWG to discuss and prepare responses to all these comments, whether or not these comments should lead to changes in the step 2 document. From this step forward, addressing the public comments and initiating work on a final guidance (step 3), the industries are participants, but the EWG is led by and final guidance signoff (step 4) is solely in the hands of the regulatory representatives.

Fig. 1.1 The formal step procedure of the ICH process for guideline development

After publishing this step 4 document by all the regulatory regions, the guideline comes into force approximately half a year after publication. Step 5 is the guidance implementation.

The ICH guidelines do not have any authority until they have been adopted by the respective authorities and published (e.g., after translation in Japan) according to the regional rules. For FDA, this is accomplished by publishing or announcing availability in Federal Register.

In the recent years of the ICH, the processes are much more efficient and structured, while in the early years of the process, it more approximated a structured scientific dialogue. As an example, the process for harmonization on assessing carcinogenic potential started in January 1992, but the final guideline (S1B) did not achieve step 4 until 1997.

Currently, ICH EWG discussions are better organized, and the SC monitors the process and progress based on initial business plan and concept paper.

1.3 Contents in ICH

Topics at the first ICH for safety were the following:

1. Toxicity testing program
2. Reproductive toxicology
3. Biotechnology

Topic 1 regarding the toxicity testing program identified several other topics that were separate issues later on in the ICH process, e.g., carcinogenicity testing (S1) and duration of chronic repeated dose toxicity studies (S4).

Also genotoxicity (S2) and toxicokinetics (S3) were identified as important topics to be started.

One issue was already solved at this conference, i.e., the requirement of the LD_{50} and the discussion on the no-adverse-effect level. The "classical" LD_{50} determination was no longer needed in any of the regions and was replaced by a well-designed, single-dose administration in a design with an increasing dose schedule with a detailed description of the pharmaceuticals' effects.

The FDA eliminated the requirement for LD_{50} testing in 1988 (Federal Register, 1988, 53 FR 39650) and published a revised acute toxicity testing guidance in 1996 that allowed use of dose-ranging studies in place of acute toxicity studies, unless the acute toxicity was to be the primary supporting data for single-dose clinical trials (Federal Register, 1996, 61 FR 43934). In the latter case, more extensive toxicological evaluation is needed than is available from routine acute studies. This guidance is essentially the guidance offered in the ICH M3 (R2) guidance but was available in the USA 15 years earlier.

A specific point in the harmonization of the rules is the translation from and into the different languages. This became especially clear with harmonization of the requirements for acute toxicity tests. Public pressure to take better care of laboratory animals led to a discussion on the necessity of the so-called LD_{50} test, a test from which many animals die. It soon became clear that an agreement could be reached by emphasizing the no-observable-effect-level (NOEL) dose, and particularly the no-observable-adverse-effect-level (NOAEL) dose.

> Language barriers cause trouble for the authorities. As an example, one Japanese term for NOAEL was translated incorrectly from Japanese into English, giving the wrong impression to Western toxicologists. "Mu Sayo Ryo" stands for no-effect dose and "Mu Dokusei Ryo" stands for no-toxic-effect dose. The difference between those two is the incorrect interpretation of the effect as "toxic" as "adverse." In an attempt to avoid confusion, the Japanese added a third term, "Mu Eikyo Ryo," meaning a dose without any biological effect, which in fact was the same as one of the earlier terms, thus causing even more confusion. Harmonization at this point was primarily a case of correct translation (Hayashi 1991).

The topic reproductive toxicology, discussed extensively in Brussels in 1991, was numbered S5. In Japan, the primary evaluation of reproductive toxicology was done on an administrative level without specific toxicological knowledge. The exposure of rats and rabbits during pregnancy will be during a standard period, e.g., from day 5 to day 16, but this can differ for various authorities, e.g., from day 6 to day 18. As Japan was especially strict in applying this rule, sometimes studies had to be repeated because the Japanese requirements were not met. It has now been agreed that small differences in intervals will not lead to a request for a new study (Brussels, abstracts).

At ICH 1, the special approach of how to assess biotechnology-derived proteins was already discussed and this topic was identified as S6. This topic, in part because

Table 1.3 ICH safety guidelines

S1	Carcinogenicity testing (3 guidelines 1995–1999) (new process just started)
S2	Genotoxicity testing (2 guidelines 1995–1997)
S3	Toxicokinetics (2 guidelines)
S4	Duration chronic studies (book chapter and guideline)
S5	Reproductive toxicity (2 guidelines, 1994)
S6	Preclinical testing of biotech-derived proteins (guideline 1997, addendum 2011)
S7	Safety pharmacology (2 guidelines)
S8	Immunotoxicity (guideline 2005)
S9	Testing of anticancer drugs (guideline 2010)
S10	Photosafety (just started)
	Multidisciplinary
M3	Timing of nonclinical studies (guideline and revision)
M4	Common technical dossier
M7	Genotoxic impurities

of the relatively new nature of this pharmaceutical field and the broad scope of the therapeutic modalities with very different considerations, took an extended period before finalization could be reached. The character of the S6 is indeed more considered and case by case than other safety guidelines, which tend to be more directive (Table 1.3).

1.3.1 The Common Technical Dossier

The topic M4 needs some special discussion, as this is rather an administrative than a scientific guideline, defining the common technical dossier. This issue was mainly driven and led by the pharmaceutical industry, as it was their special interest to develop a unified submission format globally. Rearranging all the studies from one system (e.g., US) to another (e.g., EU) was an enormous task including renumbering the studies and rewriting the cross-references. This task could take another 3 months with no other purpose than making the same scientific material available in another part of the world in a differently specified format.

With respect to the European dossier (as defined in Directive 75/318), this was an important change, which was, however, welcomed by the European assessors, as the sequence till then was not logical, starting with acute toxicity and having the pharmacodynamics as part F. It is now more logical in that it starts with the fundamental pharmacological properties of the compound providing an underlying basis for assessing its toxicological profile.

In the generation of the recommended overview documents, there was very little in the way harmonization achieved. In Module 2, the summaries and overviews are

now included as separate segments of the CTD. In 2.4, the company should provide a so-called nonclinical overview, which should be a critical description and discussion of the whole nonclinical dossier, describing the strategy chosen to provide evidence for the proof of concept and the safety of the active substance.

In 2.6, the company is requested to give short summarizing descriptions of all studies provided in the dossier, and these nonclinical summaries should be without any critical analysis but just a factual reflection of the data from the study.

It is clear from the description that the nonclinical overview (2.4) is in fact the European Expert Report, whereas the nonclinical summaries (2.6) are the same as the factual summaries from the former FDA requirements. Thus, "harmonization" was achieved by requiring both pre-existing requirements.

In practice, the two documents are complementary. The nonclinical overview (2.4) is often used as the start of the European assessment report. When sufficient details are lacking, the nonclinical summaries usually include the necessary additional study-specific detail.

1.3.2 Reduction, Refinement, and Replacement of Animals

The rational use of live animals to test the safety of new pharmaceutical entities was on the agenda of the ICH from the very beginning. It was, and still is, a held belief of those involved in the ICH process that distinctly different approaches to testing and study design should not be used in different regions of the world to prove the safety of a new pharmaceutical. Harmonization was explicitly meant to reduce the redundancy of animal experiments, without reduction of the safe development of pharmaceuticals. Some efforts in this regard have been mentioned above.

As examples, we can mention the following achievements within the framework of ICH in this respect:

- A better definition of what is needed as the purpose from acute toxicity studies. Initially, this lead to a reduction of redundancy, but recently with the revision of ICH M3 elimination of the requirement for acute toxicity studies, as the most valuable information from such studies can be derived from an appropriate observation in a repeated dose toxicity study.
- A more flexible interpretation of the interval of exposure in embryo–fetal toxicity tests, which reduced the repetition of this type of study.
- Reduction of the chronic rat study requirements, i.e., only one study of 6 months duration. The requirement for a 12-month study was eliminated as for this type of products (intended for long-term duration of therapy) a carcinogenicity study of 24 months will be generally conducted, and the primary new information from 12 month studies was relevant to tumor observations.

- Refinement (better welfare conditions of animals) can be found in the guideline on dose selection S1C. The maximum dose to be applied is no longer only the maximum tolerated dose, i.e., associated with toxicity, but can be based also on other criteria, such as a 25-fold AUC exposure as compared with the intended human exposure at therapeutic levels.

Not all ICH guidelines developed, however, have reduced animal use. Since issuing the guideline on toxicokinetics (ICH S3), an increase in use of animals was noted, as extra animals were needed for sampling during the test period so as not to compromise the toxicity observations, particularly in rodents.

1.3.3 Assessing Toxicity Versus Assuring Safety at Clinical Level

The contrast in perspectives of the ICH EWG parties, first noted in the S1 guidance discussions, was often referred to in shorthand as the US FDA wanting "toxicity studies," the EU and Japanese authorities satisfied with "safety studies." The industry then was perceived to be focused on conducting "pharmacodynamic studies" to assess human risk. With such different views on a fundamental objective of toxicology study design, it is no wonder that significant conflict existed in interpretation of study adequacy. While no longer quite as polarized as in the early 1990s when the carcinogenicity study dose selection discussion took place, hints of this same fundamental distinction in philosophies can be seen even in ICH guidelines discussions ongoing nearly 20 years later. One needs to look no further than the recently revised ICH M3 (R2) guidance (2009). It endorses high dose selection criteria for general toxicity studies that include a 50-fold exposure multiple of the clinical therapeutic exposure. For this endpoint, however, there are still clear regional differences in its ultimate acceptability. The FDA requires demonstration of dose-limiting toxicity in one species at some point in development regardless of the 50-fold multiple being achieved without toxicity. In contrast, other regulatory regions accept the exposure limit without ever requiring demonstration of toxicity, a vestige of this "toxicity-based" mindset. In recognition of the progress in understanding that has occurred during the intervening time, the viewpoints have migrated toward a more middle ground. High-exposure multiples are accepted as dose selection endpoints in many different circumstances of toxicology testing that have been written into ICH regulatory guidance. However, the different philosophies first made evident in the discussion on carcinogenicity dose selection still remain. Unfortunately, the need to bridge these underlying philosophical differences was not and often is not immediately recognized by the ICH EWG members. As a result, substantial effort has been repeatedly devoted to technical argument and proposed resolutions without consideration for the fundamental drivers of the different regional guidance and the industry's viewpoint over the course of developing many of the currently available guidances.

1.3.4 Regulatory Protocol Approval

When discussing the S1 documents on carcinogenicity testing, another type of disharmony became evident between the various regions, i.e., protocol approval. It was intimated that before initiation of the carcinogenicity study, an industry sponsor would be advised to assess the acceptability of the dose selection endpoint being proposed and that in the United States, it was considered advisable to do so by consulting with the FDA. This was incorporated in the original S1C dose selection step 2 as a note to the document. In the discussion of the Pharmacodynamic Endpoints and Additional Endpoints, a concern was highlighted that some endpoints may not be acceptable in some specific applications, and thus, consultation was warranted. This recommendation in Note 10 was deleted in later versions of the guidance based on several considerations. First, it was always appropriate to consider whether an endpoint selected was scientifically sound. Second, the EU, PMDA (and MHLW) and industry considered the correct dose selection the responsibility of the sponsoring company, not necessarily the obligation of the regulatory authorities. Third, the specific reference to and opportunity for consultation with the FDA was without an equivalent opportunity for consultation with other regulatory authorities and at the time appeared to give pre-eminence to the FDA. While the recommendation was removed from the ICH guidelines in its final form, a specific FDA guidance (FDA 2002) was created that institutionalized the practice of consultation on carcinogenicity study protocols and was included as an action in the FDA's PDUFA protocol review obligations. This recommendation was, thus, the initiation of the practice still ongoing today wherein the FDA's Carcinogenicity Assessment Committee reviews and provides recommendations for carcinogenicity study design and dose selection. This FDA practice was later expanded to include evaluation of the appropriateness of using transgenic mice for carcinogenicity studies as allowed in S1B and discussed later.

Although this paragraph focuses mainly only on carcinogenicity studies, it is still true that by the current IND process for clinical trials in the USA, the FDA can exert an enormous influence compared to other regulatory regions in directing the development of a product, by requesting additional toxicity studies or advising on approaches during the early development phase. In Japan, such an intervention system does not exist, and in fact an early assessment of nonclinical data, e.g., before first entry into humans, is not possible there. Only a registration procedure for clinical trials is sufficient to inform the authorities that such studies will start. In Europe, the clinical trial approval is not organized centrally, as it is for an important part of the marketing authorization of products, but is left at the national level. In some countries, early development review is primarily left to the local Medical Ethical Committees. It is, however, possible to request scientific advice from the various national authorities as well as from the European Medicines Agency.

In 2011 an initiative was created between FDA and EMA in the field of oncology drugs to have teleconferences exchanging information and views on actual requests

	Topic	Process
S2A/B	Genotoxic potential	Revision
M3	Timing of nonclinical studies in relation to clinical trials	Revision
S9	Nonclinical requirements for anticancer drugs	New topic
S6	Safety assessment of biotechnology-derived products	Revision

from industry harmonizing the approach between the two authorities with regard to safety for this product class.

1.4 Restart of Safety Guidelines

It was after completion of the Guideline on Immunotoxicity testing (S8) in 2005 that a need for revision of some of the safety guidelines was identified. In addition, some new topics were also to be considered. In a brainstorming session conducted in June 2006, representatives of all ICH parties, observers, and interested parties came together in Yokohama. The following topics were identified:

Based on a priority voting by the group, it was agreed to start with the revision of the S2 guidelines and of the M3 guideline. Also, issuing a guideline on nonclinical requirements for anticancer drugs was felt to be urgent and was accepted as a new topic

Although revision of the S6 safety assessment of biotechnology-derived products was found to be more urgent than the issue of anticancer drugs, a period was agreed upon to organize regional discussions to identify the topics that had to be updated in this document. These meetings were held in 2007 in the various regions, e.g., at the Japanese Drug Evaluation Forum in August 2007 and during the Immunotoxicity Summer School in Lyon in France in October 2007. During the ICH Steering Committee meeting with EWG's in Portland 2008, it was decided which topics were to be taken on board.

After finalization of the revisions of S2 and M3, there was an opportunity to take on the work of new topics, i.e., on photosafety testing S10, a new multidisciplinary topic M7 on genotoxic impurities. As these guidelines are just in the draft stage (June 2012), no descriptions are included in this book. In June 2012 also a process started to reconsider the strategy to assess the carcinogenic potential. In the chapter on S1, future perspectives, there are highlights of some of the relevant considerations.

1.4.1 Harmonization of Contents Versus Harmonization in Systems

Harmonization of regulatory guidelines does not imply that regulatory authorities are harmonized in their way of organization, approaches or processes. This can be nicely illustrated by the following example.

During the brainstorm session in 2006, mentioned above, the safety EWG, working on a revision of the documents from the first round, concluded that for S1C a "quick-win" could be reached by omitting a single word, i.e., "non-genotoxic" (and a note referenced in that sentence), from the guidance and would expand the utility of the guidance. All parties were supportive of this solution, and it was accepted immediately by all six ICH parties. Unfortunately, this did not consider the requirements for FDA legal review of all guidance documents. Once opened, regardless of whether text had been changed and which was established as acceptable ten years previously, a new legal review was required. According to the legal review, now, even the word "acceptable" was no longer acceptable for use in guidance and needed (also not an acceptable word) to be changed to "appropriate." The Japanese expert previously involved in developing the guidance stated, however, that at some points in the text, the word "acceptable" was more appropriate than the word "appropriate," and so he deemed the substitution "unacceptable." Considerable time then had to be spent to find wording that was "appropriate" as well as "acceptable."

In fact this specific discussion led to generating a list of words that were not considered appropriate for use in ICH guidelines document, according to FDA legal review.

We can learn from this experience that the legal culture of an organization is a part of the business that we have to accept and that thought tipped the scales in coming to a conclusion together. You cannot harmonize everything. You have to bear in mind that you are dealing with regulations for medicinal products, but you cannot harmonize all the medical practices. That is true within an ICH region and for the ICH process as a whole.

1.4.2 Development of Regulatory Pharmaceutical Toxicological Sciences

The ICH process on safety has stimulated to a great extent a specific type of applied science, i.e., regulatory aspects underpinning the guidelines written on behalf of the authorities responsible for the assessment of human medicines. (It also brought to life a similar process for veterinarian therapeutics.) Although probably not unique for the field of human pharmaceuticals, the existence of a scientific procedure as in ICH has led to an important stimulation of conducting comparative research in the datasets derived from the pharmaceutical dossiers, within the pharmaceutical trade associates, the regulatory authorities and associated academia. Nearly all chapters in this book illustrate this in that data evaluations drove or underpinned most of the critical recommendations. Some processes are preceded by various publications; others have led to parallel research, especially in the early days. An important contributor has to be mentioned in this case: Dr. Joseph Contrera from the FDA Office of Research. His paper on the comparison between outcomes of 6 vs. 12 month studies in non-rodents (Contrera et al. 1993) is not undebated (see chapter on S4) but despite that is very important. His contributions in the processes of S1 with regard to exposure comparison (Contrera et al. 1995) were very important to understand the issues of dose selection

in carcinogenicity studies. His database on carcinogenicity studies (Contrera et al. 1997) published in 1997 was based upon 282 carcinogenicity studies on pharmaceuticals only, a very large number of studies, which was not equaled even today.

1.5 Final Conclusion

Overseeing the ICH process after more than 20 years can teach us that an in-depth mutual understanding has been reached between the various regions and also between pharmaceutical industry and regulators. In this chapter and this book, we have focused ourselves on the toxicological aspects, but it is true too for other areas which are critical in the scientific criteria to authorize a medicinal product, i.e., quality and efficacy.

The frequency of discussions within ICH is currently less than at the start of the process 20 years ago, for the obvious reason that the main topics have been discussed already but also due to resource constraints. It is, however, to be expected that there are significant accomplishments yet to be realized in rationalizing, simplifying and improving pharmaceutical development and that the ICH has a significant leadership role to play in this future.

References

Cone M (2001) Proceedings of the fifth international conference on harmonisation, San Diego 2000. The Regulatory Press, 698p

Contrera JF, Aub D, Barbehenn E, Belair E, Chen C, Evoniuk G, Mainigi K, Mielach F, Sancilio L (1993) A retrospective comparison of the results of 6 and 12 months non-rodent toxicity studies. Adverse Drug React Toxicol Rev 12(1):63–76

Contrera JF, Jacobs AC, Prasanna HR, Mehta M, Schmidt WJ, DeGeorge JJ (1995) A systemic exposure-based alternative to the maximum tolerated dose for carcinogenicity studies of human therapeutics. J Am Coll Toxicol 14:1–10

Contrera JF, Jacobs AC, DeGeorge JJ (1997) Carcinogenicity testing and the evaluation of regulatory requirements for pharmaceuticals. Regul Toxicol Pharmacol 25:130–145

D'Arcy PF, Harron DWG (1992) Proceedings of the first international conference on harmonisation, Brussels, 1991. Queen's University, Belfast, 590p

D'Arcy PF, Harron DWG (1994) Proceedings of the second international conference on harmonisation, Orlando 1993. Queen's University, Belfast, 736p

D'Arcy PF, Harron DWG (1996) Proceedings of the third international conference on harmonisation, Yokohama 1995. Queen's University, Belfast, 998p

D'Arcy PF, Harron DWG (1998) Proceedings of the fourth international conference on harmonisation, Brussels, 1997. Queen's University, Belfast, 1158p

Food and Drug Administration. Carcinogenicity study protocol submissions. http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm065014.htm. Accessed 22 Feb 2002

Hayashi Y (1991) No effect dose vs no toxic effect dose and post dosing reversibility/'delayed toxicity'. In: D'Arcy PF, Harron DWG (eds) Proceedings of the second international conference on harmonisation, Brussels, 1991. Queen's University, Belfast, pp 191–193

Chapter 2
EU Perspective on ICH

Spiros Vamvakas

Abstract This chapter is considering ICH I in the context of the EU regulatory perspective, starting with a history of ICH in 1989, a time when the EU was pioneering a single pharmaceutical market in the EU. One major achievement of ICH, the agreed Common Technical Document for regulatory submission, is described in detail. Furthermore, the chapter explains how the ICH guidelines are implemented by the European Medicines Agency in the EU regulatory system. Given the fact that ICH has already a 20-year history, this chapter also elaborates on how important it is to maintain the guidelines, once adopted, by revising them or complementing them with addendums and/or questions and answers document updates based on new science or to ensure harmonised implementation. Finally, the chapter describes the efforts of ICH to provide training to developing countries, newly instituting their own pharmaceutical regulations and guidance, and to reach out beyond the EU, Japan and the USA and encompass new regions which have become important in drug development since the formation of ICH.

2.1 Introduction

The International Conference of Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH), launched 20 years ago, brings together the drug regulatory authorities of Europe, Japan and the USA, along with the pharmaceutical trade associations from these three regions, to discuss scientific and technical requirements for the development of medicinal products.

ICH's goal is to achieve greater harmonization in the requirements for product registration, thereby reducing duplication of testing and reporting during the research and development of new medicines.

S. Vamvakas (✉)
European Medicines Agency, London, UK
e-mail: Spiros.vamvakas@ema.europa.eu

J.W. van der Laan and J.J. DeGeorge (eds.), *Global Approach in Safety Testing*, AAPS Advances in the Pharmaceutical Sciences Series 5, DOI 10.1007/978-1-4614-5950-7_2,

2.2 Brief History

Harmonization of regulatory requirements for medicinal products was pioneered by the European Community in the 1980s as the EU moved towards a single market for pharmaceuticals. Since the initiation of this effort, EU regulators have achieved what appeared almost impossible a few decades ago, a harmonised market across the 27 EU member states (Fig. 2.1).

In the WHO Conference of Drug Regulatory Authorities (ICDRA) in Paris in 1989, plans for harmonization among the leading regions for new pharmaceuticals EU, USA and Japan began to materialise. In April 1990, in a meeting hosted by EFPIA in Brussels, the ICH Steering Committee (SC) was established, and the first meeting of the ICH Steering Committee took place in October 1990 in Tokyo. In addition to the Steering Committee, the ICH structure encompasses the ICH coordinators, the ICH secretariat and last but not least the ICH Expert Working Groups, which develop and maintain the guidelines which are then adopted by the SC. All participants meet face-to-face at least twice per year and work collaboratively in the periods in between remotely.

Fig. 2.1 EU member states have a harmonised market for pharmaceuticals

The guideline development encompasses several steps. The process is kicked off by one or more ICH members proposing a new topic with a concept paper and business case to justify why this new topic will contribute to harmonization of requirements for registration and the expected benefits. Once the topic is accepted by the Steering Committee, the experts develop a draft guideline. After adoption of the draft by the SC, this "Step 2" document is published in each region for broad stakeholder consultation locally. At the end of the consultation, the Expert Working Group reconvenes to discuss the comments and prepare the final guideline, "Step 4", which comes into force in general 6 months later (Step 5 once implemented).

ICH has published and continues to maintain more than 50 guidelines on:

Q: quality, e.g. stability, analytical validation, impurities, pharmacopoeia harmonization, quality of biotechnological products, specifications, good manufacturing practice, pharmaceutical development, quality risk management, quality systems and chemical/biotechnology common guideline on the active substance
S: non-clinical safety testing, e.g. toxicity, carcinogenicity and genotoxicity studies, toxicokinetics and pharmacokinetics, reproductive toxicology, special aspects in toxicity testing of biotechnological products, pharmacology studies, immunotoxicology studies, safety of oncology products and photosafety
E: clinical efficacy and safety, e.g. clinical study reports, dose–response studies, ethnic factors, good clinical practice, general guidance on clinical trials, statistics, paediatrics, clinical safety, electronic submission of case safety reports, geriatrics, QT prolongation, pharmacogenomics definitions and data submission and development safety update report
M: multidisciplinary topics, e.g. medical dictionary for regulatory activities terminology/MedDRA, data elements and standards for drug dictionaries, and preclinical trials in relation to clinical trials

Last but not least, ICH is credited with the development of the Common Technical Document (CTD) and its electronic form (eCTD), a critical communication tool supporting the registration of new pharmaceuticals across the ICH regions.

2.3 Common Technical Document

One major achievement of ICH is the development of the Common Technical Document (CTD) which revolutionised the submission procedures for industry. The creation of this single technical dossier and later its electronic form, the eCTD, accepted by all the three ICH regions, resulted in significant savings in time and resources, facilitating simultaneous submission, review and approval of new drugs. Prior to the CTD, industry spent irrational amounts of time trying to adjust technical data formats to the specified formats of the different regions. The format in the EU reflected at that time the format as required by the EU Directive 75/318, while other regions had other formats. The CTD not only greatly accelerated the preparation of marketing authorisation applications for industry but also made the exchange of

information among drug regulatory authorities easier, facilitating discussions of important topics during the evaluation of applications.

The CTD was a major step forwards because it also enabled the creation of the electronic format of submission, the eCTD, which transformed the marketing authorisation application from many thousands of pages in numerous boxes to be delivered to the Agencies to paperless delivery in electronic format. But it is not only the delivery which was made easier with the introduction of the eCTD; also the review was greatly facilitated by the new, easily navigatable format which enabled the exchange of inquiry and response of the pharmaceutical review and evaluation process.

2.4 ICH and the EU Regulatory System

In the EU, pharmaceutical guidelines can be grouped either as regulatory or scientific.

The basic EU legislation is supported by a series of regulatory guidelines published by the European Commission. A regulatory guideline is a document with explicit legal basis referred to in the legislation and intended to provide guidance to industry, regulators and/or other interested parties on the best way to fulfil a legal obligation.

Scientific guidelines are intended to provide a basis for practical harmonization of the requirements of the European Medicines Agency (EMA) for the demonstration of quality, non-clinical pharmacology and toxicology (safety) and efficacy (investigation of clinical efficacy and side effects) for new medicinal products. Scientific guidelines which cover a range of topics across quality, safety and efficacy are called multidisciplinary (see above).

Scientific guidelines also help facilitate the preparation of applications for marketing authorisation by the pharmaceutical industry.

ICH guidelines are normally part of the scientific guidelines adopted by the Committee for Medicinal Products for Human Use (CHMP). In terms of clinical development, ICH covers the general requirements, while the requirements for specific therapeutic fields are covered by regional guidelines of the ICH regulators. In the EU, the CHMP develops the guidelines relating to investigation of medicinal product in specific therapeutic fields, e.g. cancer, diabetes, schizophrenia, etc.

In addition, some ICH guidelines have been integrated into EU legislation. For example, following the adoption of the ICH guideline Q7 (good manufacturing practice (GMP) for active pharmaceutical ingredients) and the E6 guideline on good clinical practice (GCP), EU legislation was amended to require GMP for starting materials and GCP for clinical trials.

In all cases, the CHMP is involved in the ICH process early, and ICH topics under development are included in the work programme of the relevant CHMP working parties or ad hoc groups for input into the process. Once adopted by the CHMP, ICH guidelines have the same status as other EMA guidelines and replace (supersede) older existing EMA guidelines that were already available on the subjects covered.

Guidelines are generally implemented 6 months after adoption, but applicants are of course free to apply them earlier.

In addition, the EMA experts are providing proposals, as do other ICH members, for new guidelines/update (revision) of existing ones in the form of concept papers outlining the scientific rationale of the proposals and business plans outlining the expected impact of the proposal on harmonization of requirements (also in terms of savings in refining, replacing, reducing animal testing) and expected resources required for the development/ revision of the guideline. This latter aspect has become particularly important in recent years as resources in all ICH members have become more limited. A new topic or a revision of an existing topic means in practical terms the formation of an Expert Working Group, which usually has between 15 and 30 members depending on the complexity of the topic. These experts need to meet at least twice per year face-to-face for a number of years until the finalisation of the guideline, which has important financial implications. The Steering Committee will, therefore, take this into account and prioritise proposals accordingly. When choosing non-clinical topics, the impact of the new guideline/revised guideline on refining, reducing and replacing animal studies is of paramount importance for the EU.

2.5 The Importance and Tools of Maintenance of Existing Guidelines: Geriatrics and Non-clinical Guidelines

ICH guidelines aim to represent the gold standard of scientific knowledge at the time they are issued. However, in many areas, science and other changes mandate an updating of the guidelines. In general, there are three tools to update the guidelines: revision of the main body of the guideline, development of an Annex to the guideline and the development of a questions and answers document, the latter of which is usually used as an implementation guide.

This book contains a comprehensive discussion of the non-clinical guidelines which have been implemented and the history of the implementation process by various members of the subject EWGs; therefore, experience with a Clinical Efficacy Guideline, which has recently been updated at the request of EU experts, has been selected for detailed discussion here.

2.5.1 Geriatrics

The initial guideline on requirements for geriatric patients “Studies in Support for Special Populations: Geriatrics” was finalised in 1993. In this guideline, it stated among others that “Geriatric patients should be included in the Phase 3 database (and in Phase 2, at the sponsor’s option) in meaningful numbers. The geriatric subpopulation should be represented sufficiently to permit the comparison of drug response in them to that of younger patients. For drugs used in diseases not unique

to, but present in, the elderly a minimum of 100 patients would usually allow detection of clinically important differences. For drugs to treat relatively uncommon diseases, smaller numbers of the elderly would be expected. Where the disease to be treated is characteristically associated with ageing (e.g., Alzheimer's disease) it is expected that geriatric patients will constitute the major portion of the clinical database".

The minimum number of 100 patients was dictated at that time, mainly by minimum requirements in terms of detecting side effects specific to the geriatric population. The demographics of the society have changed rapidly in the years since this guideline was finalised, and new drugs are used extensively in elderly patients including those aged over 65 (the conventional definition) but also above 75 and above 85 (the real elderly population of our times) without proper knowledge of their safety and efficacy in this population.

Around 2006, the European Union Geriatric Medicines Society (EUGMS) raised the possible need for an EU "geriatrics" legislation to address the need for clinical trials in the elderly based on the rationale that there are complex changes of pharmacokinetics/pharmacodynamics (PK/PD), due to ageing, co-morbidity and polypharmacy and that the efficacy and safety of drugs in older people cannot be deduced from randomised clinical trials performed in young and adult subjects or from meta-analysis including a small number of subjects. There was intense discussion in the EU and internationally on this issue, and as an alternative more flexible and more global proposal, a revision of the ICH guideline on geriatrics was tabled. ICH regulators reviewed the geriatric data in marketing application submissions, and the conclusion was that the vast majority of applications had 100 geriatric patients, not less but also not more, which was no longer considered acceptable.

In the Steering Committee meeting in Yokohama in November 2007, the EU presented a CHMP concept paper proposing a revision of the geriatrics guideline to reconsider age cut-offs, the very elderly, frail elderly; co-morbidities; PK/PD interactions; specific PK studies; and specific formulations. The Steering Committee adopted the EU proposal to convene an informal expert group to work via teleconference with a view of preparing a proposal for the next meeting. The EU was appointed rapporteur, and the proposal was adopted in June 2008 in Portland. The expert group was mandated with drafting a questions and answers document to better reflect the current requirements in this age group.

The Q&A document adopted in September 2010 changed the previous approach. While maintaining the flexibility of the initial document, the new document emphasised the need too have sufficient data in the populations reflected in the demographics of the disease to assess the benefit/risk in these populations:

> Geriatric patients can respond differently from younger patients to drug therapy in a number of ways and such differences can be greater in patients 75 years and older:
>
> (a) The geriatric population has age-related physiological changes that can affect the pharmacokinetics of the drug, and the pharmacodynamic response to the drug, both of

> which can influence the drug-response and the dose response relationship. (b) Geriatric patients are more prone to adverse effects since they often have co-morbidities and are taking concomitant therapies that could interact with the investigational drug. The adverse effects can be more severe, or less tolerated, and have more serious consequences than in the non-geriatric population. With the increasing size of the geriatric population (including patients 75 and older) and in view of the recent advances in pharmacokinetics and pharmacodynamics since the ICH E7 guideline was established in 1993, the importance of geriatric data (from the entire spectrum of the geriatric patient population) in a drug evaluation program has increased." … "In the marketing application, depending on the numbers of patients, data should be presented for various age groups (for example <65, 65–74, 75–84 and >85) to assess the consistency of the treatment effect and safety profile in these patients with the non-geriatric patient population. As single trials may not have sufficient numbers of geriatric patients to allow such analyses, these will often need to be carried out on pooled data. Any such analyses will need to consider consistency across studies.

This new approach to geriatrics was the goal of the EU regulators when they proposed to revisit this guideline.

2.5.2 Non-clinical Guidelines

In the spring of 2006, EU regulators proposal on the review of ICH Safety (non-clinical) Guidelines was circulated to the ICH Steering Committee. The proposal was justified on the basis of better regulation and the need to keep guidelines up-to-date and focused on implementation of guidelines, as well as the high political importance of ensuring that the use of animals in drug development is kept under review in the context of the 3R Agenda: refinement, reduction and replacement of animal experiments. The EU team reported on the work carried out by the CHMP Safety Working Party to review all ICH guidelines. The EU non-clinical experts recommended a review of the S2 genotoxicity guidelines, the S6 guideline on preclinical safety for biotechnology products and the M3 guideline on timing of non-clinical studies. The Steering Committee accepted the EU proposal for the organisation of an informal meeting of experts in Yokohama in June 2006 to discuss the need for a review of ICH non-clinical safety guidelines and make recommendations to the SC based on these discussions.

This EU proposal resulted in major revisions to the harmonised requirements in all three areas to reflect the current state of the art. The process also showed that revising a guideline is at least as difficult and time consuming as drafting a new guideline also due to the fact that adoption of changes in established approaches by all six ICH parties is a very difficult task. The addendum to the S6 guideline was finalised in June 2011, the revision of the M3 guideline was finalised in June 2009 and the related questions and answers document in June 2011 and finally the revision of the S2 guideline was finalised in November 2011.

2.6 ICH Reaching Out to the World Beyond: The Global Cooperation Group

For the first 10 years or so, ICH focused on the development of guidelines and standards for use in the ICH regions, i.e. European Union, Japan and the United States. By the late 1990s, however, ICH recognised the growing interest in ICH guidelines beyond the ICH regions. On the one hand, there was a growing recognition of the broader utility of ICH guidelines. On the other, the globalisation of industry drove a need for common standards both in ICH and non-ICH regions with significant role in the development and utilisation of new drugs.

This was the basis for the creation of the Global Cooperation Group (GCG) in 1999. The goal was better understanding of ICH guidelines through open communication and dissemination of information facilitated by trainings.

From the beginning, it was made clear that that GCG does not aim to impose ICH guidelines on any country or region and that the GCG will work closely with WHO and other international organisations to achieve harmonization and greater utilisation of ICH guidelines.

Partnerships were created with Regional Harmonization Initiatives (RHI), networking national authorities in all parts of the world such as the Asia-Pacific Economic Cooperation (APEC), the Association of the Southeast Asian Nations (ASEAN), the Gulf Cooperation Council (GCC), the Pan American Network for Drug Regulatory Harmonization (PANDRH) and the Southern African Development Community (SADC).

Training in the broad sense is a key GCG focus and the EU experts have been very active in delivering training in many non-ICH regions in the last years. But as recent workshops on clinical trial assessment and inspection showed, training has moved beyond simply an understanding of ICH guidelines to the active consideration of application of ICH guidelines in the assessment of studies and data.

2.7 The Globalisation of the Pharmaceutical Market and the Regulators' Forum

More recently, ICH recognised the need for further change to mirror the global face of drug development. This led to the creation of the Regulators' Forum in 2007 to enable the representation of individual drug regulatory authorities (DRAs) from regions that were either a major source of active pharmaceutical ingredients (APIs), clinical trial data, or had adopted ICH guidelines. The participation of DRAs is distinct but also complementary to that of Regional Harmonization Initiatives representatives in the GCG.

The first forum took place in 2008 in Portland. Regulators were invited from countries with a history of ICH guideline implementation (Australia, Chinese Taipei, Singapore and South Korea) and also from countries which are currently

important in manufacturing of medicinal products and contacting clinical trials, such as China, India, Brazil and Russia.

Compared to GCG, the focus of the Regulators' Forum is to create a regulator-only environment for open discussion of issues related to the *implementation* of ICH guidelines for regulators around the world. In the meanwhile, the Regulators' Forum has established itself as a very useful satellite meeting of every ICH meeting and has succeeded in facilitating communication and interactive contact among ICH and non-ICH regulators with topics often around similarities and differences in the interpretation of ICH guidelines across regions. Some non-ICH countries, such as Australia, opted to harmonise their own requirements by adopting what were then seen as international best practice standards, and they chose the ICH guidelines as benchmark. A factor in those decisions was the emerging reality: the pharmaceutical industry was increasingly globalised, and the regulatory requirements for new and innovative medicines were best reflected in the developing ICH guidelines which at that time represented all major regions in terms of drug manufacturing and non-clinical and clinical research.

2.8 Outlook

ICH recognises that the world has changed since its creation, and new regions have become important in drug development in addition to the original members EU, USA and Japan. This is, however, not a reason to discontinue ICH. ICH should be used as a very successful international platform with a measurable significant output to link all players together for the benefit of drug development.

Chapter 3
The Value and Benefits of the International Conference on Harmonisation (ICH) to Drug Regulatory Authorities: Advancing Harmonization for Better Public Health

Justina A. Molzon

Abstract Globalization of the pharmaceutical industry has created the need to harmonize the regulatory requirements for the development of new pharmaceuticals. Experts from the pharmaceutical industry and regulators joined together to establish the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) resulting in over 60 harmonized guidelines. After agreement had been reached on these guidelines, the next logical step was to arrange the information in a common format for submission. To this end, ICH has established the Common Technical Document (CTD) and the electronic Common Technical Document (eCTD). The practical issues of implementing ICH Guidelines will be presented. Further, the value and benefit of ICH to regulators and the evolution of the CTD/eCTD as a common regulatory language will be detailed.

3.1 Value and Benefits of the Common Technical Document

The International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH), launched over 20 years ago, is an unparalleled undertaking. ICH brings together the drug regulatory authorities of Europe, Japan, and the United States, along with the pharmaceutical trade associations from these three regions, to discuss scientific and technical aspects of product registration. It is ICH's mission to achieve greater harmonization in the interpretation and application of technical guidelines and requirements for product registration, thereby reducing duplication of testing and reporting carried out during the research and development of new medicines.

J.A. Molzon, M.S. Pharm., J.D. (✉)
Center for Drug Evaluation and Research, U.S. Food and Drug Administration, Silver Spring, MD, USA
e-mail: Justina.Molzon@fda.hhs.gov

J.W. van der Laan and J.J. DeGeorge (eds.), *Global Approach in Safety Testing*, AAPS Advances in the Pharmaceutical Sciences Series 5, DOI 10.1007/978-1-4614-5950-7_3,

In 2000, the 10th Anniversary of ICH, Dr. Caroline Nutley Loew of the Pharmaceutical Research and Manufacturers of America (PhRMA) wrote a report, *The Value and Benefits of ICH to Industry*, which detailed ICH's creation, procedures, and guideline development in the areas of safety, efficacy, and quality. Dr. Loew's report anticipated that the Common Technical Document (CTD) would revolutionize the submission procedures for industry's regulatory staff. Dr. Loew characterized the CTD as "offering potential benefits to industry far greater than any other single ICH topic" and predicted the CTD would afford significant savings in time and resources as complex multiple submissions were replaced by a single technical dossier submitted in the three ICH regions—facilitating simultaneous submission, approval, and launch of new drugs. In calling the CTD "a topic whose value to industry cannot be underestimated," Dr. Loew noted that with full incorporation of the CTD and the electronic CTD (eCTD), ICH could turn its sights to disseminating guideline information to non-ICH countries, yielding additional benefits to both regulators and industry.

Ten years later and in anticipation of ICH's 20th Anniversary, the value and benefits of ICH to regulators have been realized. Moreover, implementation of the CTD in 2003 promoted the involvement of drug regulatory authorities (DRAs) not initially part of ICH, thereby extending ICH's harmonized approach. The development of the Global Cooperation Group, which includes representatives from five regional harmonization initiatives and the newly established Regulators Forum, created to promote participation by non-ICH countries interested in implementing ICH's strategies, have also helped incorporate the CTD into regulatory processes, creating a common regulatory language that promotes faster access to life-saving treatments to patients beyond ICH regions. In recognition of the increasingly global face of drug development, ICH recently updated its logo to emphasize the benefits of harmonization for better global health.

3.2 Shift in Emphasis

Substantial benefits to DRAs resulted when ICH shifted emphasis from the input of information by industry to the output of information by regulators. This transition was made possible by the development of a common submission format—the CTD—which greatly influenced regulatory review processes, ultimately leading to a harmonized electronic submission and e-review initiatives, which, in turn, have enabled implementation of good review practices. These activities are having a global effect on information review and sharing among drug regulatory authorities.

Originally, ICH focused on input by industry—the technical submission requirements for pharmaceuticals for human use. Harmonizing the differences in these requirements through ICH Guidelines helped industry reduce development times and save resources. To extend the benefits of harmonization, industry proposed assembling the building blocks of information intended for inclusion in a submission into a consistent harmonized format,

referred to as the CTD, which would relieve pharmaceutical companies of the time, workforce, and financial burdens of assembling a submission for one DRA and then having to reformat it for another. This new consistent format also greatly benefited the U.S. Food and Drug Administration (FDA), enabling the agency to establish templates for each of the review disciplines while promoting more consistent review practices and processes.

Prior to the advent of the CTD, regulatory reviewers received an application from one company and spent a year or more engaged in its review. When the review was completed, reviewers received the next application—most likely in a different format—and had to learn the structure of the new application. As a result, review staff were constantly on a learning curve when new assignments were received—time they could have better used reviewing the information as opposed to simply trying to find it.

When industry proposed the CTD in 1996, ICH regulators were hesitant to change their submission formats, believing it would be too disruptive to the review process. They needed convincing that harmonizing the submission format had value. Regulators asked industry to do a feasibility study. That study, conducted in May 1996, evaluated the time it took to convert an FDA new drug application into a European Medicines Agency (EMA) submission and the reverse. It also evaluated the number and types of staff needed to carry out the conversion of the submission formats. Regulators quickly saw the potential value of harmonizing submission formats.

The CTD has also made the exchange of information among drug regulatory authorities easier. For a number of years, FDA and the EMA have had a confidentiality arrangement in place allowing the sharing of confidential information, greatly increasing interactions between the two agencies. Now that submissions are received in the same format, and, generally, at the same time, these interactions have become more efficient, facilitating discussions of common concerns as submissions are evaluated.

Last, and perhaps most important, the CTD has facilitated electronic submissions (the eCTD). In the past, drug applications were voluminous, delivered to FDA by the truckload, due to the sheer amount of paper involved. When the agency first transitioned to electronic submissions, an application was on a compact disc or hard drive. Although this certainly helped with transportation and storage issues, it did not necessarily enhance the review process. FDA has now implemented the FDA Electronic Submission Gateway, which allows a new drug application (NDA) to be sent electronically, essentially very much like e-mail. After being assessed for completeness, a submission is immediately and fully accessible on the reviewer's desktop. This innovation has alleviated the need for industry to create and assemble the many pieces of paper that constituted a traditional paper-based product application, organize the application, box thousands of pages, load the boxes on a truck, and deliver them to FDA—all before a reviewer could even begin the assessment process.

The eCTD has proved critical to improving application submission efficiencies as well as reviewer efficiency. Besides delivering submission material to the reviewer in an expedited manner, the eCTD format has made it easier to develop standardized reviewer e-templates and review tools for each of the review disciplines.

Another benefit of a harmonized format has been the ease of developing and implementing harmonized good review practices. What is evaluated in a review is closely tied to the requested data. As a result, there is considerable similarity between ICH Guidance to industry and what we consider good review practices. Because ICH regions have harmonized much of the information submitted for marketing authorization, ICH regulators could easily begin moving toward similar review practices.

In general, good review practices promote transparency and consistency, both of which are very important if industry and the public are to understand how regulatory authorities carry out their responsibilities. This is especially important because of the complexity of the disciplines and specialties involved in the review process. We needed a consistent approach to evaluating submissions and reaching conclusions, and the CTD and eCTD have helped to achieve these goals.

In summary, the CTD format influences the content of the review by imposing a consistent order of information and data. This shapes both the conduct of the review and the presentation of the results of the review and promotes good review practices and increased efficiencies. As more countries embrace ICH Guidelines and the CTD format, a common regulatory language could evolve that will further promote interactions among drug regulatory authorities.

As previously mentioned, each ICH region implements ICH Guidelines according to its own rules and regulations. For the FDA, this means complying with GGP regulations. This should not be interpreted as undermining ICH Guidelines in anyway. GGPs help make the guidance development process as transparent as possible. ICH Guidelines still represent the Agency's current thinking on the scientific and technical information being submitted for the registration of pharmaceuticals for human use. ICH continues to be an important initiative for FDA and helps ensure that safe, effective, and high-quality medicines are developed and registered in the most resource-efficient manner.

U.S. FDA's Implementation of ICH Guidelines

Upon completion of the ICH process for a harmonized guideline, the guideline moves immediately to the final step of the process—regulatory implementation. This step is carried out according to the same national/regional procedures that apply to other regional regulatory guidelines and requirements, in the European Union, Japan, and the United States.

In the United States, the finalized ICH Guideline is implemented according to Good Guidance Practices (GGPs). GGPs are FDA's policies and procedures for developing, issuing, and using guidance documents. [*Federal Register*: September 19, 2000 (Volume 65, Number 182) Page 56468–56480]. This is why the ICH Guidelines posted by FDA are called Guidance and reformatted to comply with GGPs.

According to GGPs, FDA Guidance may be categorized as Level 1 or Level 2.

(continued)

Level 1 guidance documents include guidance documents that:

1. Set forth initial interpretations of statutory or regulatory requirements
2. Set forth changes in interpretation or policy that are of more than a minor nature
3. Include complex scientific issues
4. Cover highly controversial issues

Level 2 guidance documents are guidance documents that set forth existing practices or minor changes in interpretation or policy.

In the context of ICH, Level 1 guidance are generally ICH Guidelines that go through the four-step ICH process, and Level 2 guidance are generally Q and As or addendums to established ICH Guidelines.

Another noticeable modification of FDA's publication of ICH Guidelines is that although guidance documents do not legally bind FDA, they represent the Agency's current thinking. Therefore, FDA employees may depart from guidance documents only with appropriate justification and supervisory concurrence.

To comply with GGPs, each ICH Guidance notes on each page that it "Contains Nonbinding Recommendations," and a box outlined in a heavy black line contains the following statement:

This guidance represents the Food and Drug Administration's (FDA's) current thinking on this topic. It does not create or confer any rights for or on any person and does not operate to bind FDA or the public. You can use an alternative approach if the approach satisfies the requirements of the applicable statutes and regulations. If you want to discuss an alternative approach, contact the FDA staff responsible for implementing this guidance. If you cannot identify the appropriate FDA staff, call the appropriate number listed on the title page of this guidance.

Chapter 4
A Japanese Perspective on Implementation of the Three Rs: Incorporating Best Scientific Practices into Regulatory Process

Yasuo Ohno

Abstract ICH has been successful to decrease inappropriate duplication of tests for candidates of pharmaceuticals. It contributed to provide harmonized guidelines respecting 3Rs principle proposed by Russel and Burtch (The principles of human experimental technique. Methuen, London, 1959). LD50 values for pharmaceuticals are not required anymore also in Japan. Circumstances where assessments of safety of metabolites are needed and that where repeated-dose tissue distribution studies are necessary were defined. All of these and other harmonization achieved through ICH contributed a lot to decrease the number of animal use and to promote welfare of animals in safety studies needed for pharmaceutical development.

4.1 Importance of Three Rs

For more economical use of human, animal, and material resources and to speed global development and to facilitate availability of new medicines whilst maintaining safeguards on quality, safety, and efficacy and to foster public health, ICH (International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use) has contributed greatly by reducing or obviating the need to duplicate the testing carried out during the research and development of new medicines. In the field of nonclinical safety evaluation of drug candidates, the ICH process has achieved harmonization of many technical guidances more than two decades from the first steering committee and Expert Working Group meeting in Brussels in 1990.

Y. Ohno (✉)
National Institute of Health Sciences, Setagaya, Tokyo, Japan
e-mail: ohno@nihs.go.jp

J.W. van der Laan and J.J. DeGeorge (eds.), *Global Approach in Safety Testing*, AAPS Advances in the Pharmaceutical Sciences Series 5, DOI 10.1007/978-1-4614-5950-7_4,

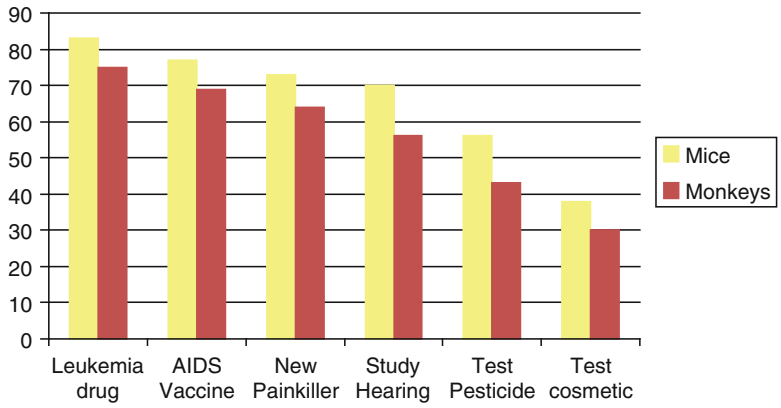


Fig. 4.1 Public attitude to animal research (Studies in UK in 1999). Prepared by Dr. Rowan (2005) from New Scientist, 22 May, 1999, 26–31.

Even though animal testing is currently essential for safety evaluation of drug candidates, there had been increasing pressure from the public to reduce or eliminate reliance on animal experimentation. On the other hand, the public attitudes have been sensitive to the need for animal experiments and allow animal experiments that were necessary for the development of pharmaceuticals, especially, for serious diseases and/or where they can be executed with less pain and distress (Fig. 4.1). Therefore, in revision of the technical guidances for nonclinical safety, it has been important to respect 3Rs principles of animal experiments (replacement, reduction, and refinement) that were proposed by Russel and Burtch (1959) in order to retain the support of the society until such time when animal experimentation can be eliminated.

In the beginning of ICH process, there were no clear expectations of contribution to the 3Rs principles. However, during the process of ICH technical discussions, it became apparent that international harmonization of guidelines and guidances on nonclinical tests were quite effective also to decrease the number of animal used for safety evaluation of drug candidates.

In general, harmonization itself decreases the unnecessary conduct of animal toxicity studies by the elimination of redundant, duplicative animal testing. In addition, improvement of the protocols for toxicity tests has also been carried out considering the 3Rs principles. The harmonized guidelines have also been written to be flexibly applied based on the information on the drug candidates and their targets, which has promoted scientific conduct of toxicity studies, thus eliminating uninformative studies. Harmonization and more efficient linkage of the timing of nonclinical testing in relation to the necessary information to support safe conduct of clinical trials and the introduction of exploratory clinical trials have helped to conduct clinical trials earlier with less animal use. Examples of the contribution of ICH to 3Rs are listed in Table 4.1.

Table 4.1 Examples of ICH contribution to 3Rs principles in animal experiments

1. Circumstances under which toxicity studies of the metabolites are needed were defined at ICH-1 (Ohno 1992) and M3 (R2) (2009)
2. Single-dose toxicity study using non-rodent does not require high dose that may cause severe toxic symptoms (Ohno 1992)
3. It became possible to replace single-dose toxicity studies in non-rodents with the other relevant data (e.g. preliminary dose-setting studies for repeated-dose toxicity studies) by the discussion in S4 (incorporated into Japanese guideline in 1993) and M3 (R2) (2009)
4. Circumstances under which repeated-dose tissue distribution studies were defined in Pharmacokinetics guidelines (S3B 1994). This reduced the conduct of the tests for most of the compounds with half-lives shorter than two times of the dosing period
5. Twelve month repeated-dose toxicity studies in rodents and non-rodents were replaced with 6 month and 9 month studies, respectively, by S4A (1998)
6. Guidelines on nonclinical safety studies for the conduct of human clinical trials indicated timing of each safety studies in relation to human clinical trials. This contributed to the drug development by less number of toxicity tests. These guidelines were M3, M3 (R1), and M3 (R2) that were harmonized in 1997, 2000, and 2009, respectively
7. Number of animal species required for usual carcinogenicity study was decreased from two to one by S1B (1999)
8. Minimum periods of repeated-dose toxicity studies before clinical trial shorter than 2 weeks in Japan and for the evaluation of toxic effects on reproductive organs were decreased from 4 weeks to 2 weeks by M3(R1) (2000) for male and M3(R2) for female (2009)
9. Introduction of exploratory clinical trials by M3(R2) in 2009 made it possible to conduct clinical tests in Japan with less animal use and earlier than before

4.2 Single-Dose Toxicity Studies and Toxicity Evaluation of Major Metabolites

4.2.1 Single-Dose Toxicity Studies

There were significant regional differences in the guidelines on single-dose toxicity studies including the number of animal species to be tested, number of the animals/group, and route of administration. To eliminate unnecessary discrepancies in the requirements of the nonclinical tests, Japanese guidelines on single-dose toxicity studies were revised in 1989 just before the first meeting of ICH in Brussels (1991). Those changes were to (1) decrease in the required number of the rodent species from two to one, (2) require only an approximate LD50 (not statistically precise LD50) values for rodents and not to require LD50 values for the non-rodents, and (3) exclude necropsies for the survived non-rodents. The reason for these revisions was the understanding that observation of the toxic features in relation to the dose is more important than the LD50 value and the value itself differs among different institutes even though the other experimental conditions are the same.

With these revisions, the differences in the requirements across the ICH regions became relatively small except for the requirement of the number of animals per groups. It was at least 5 in rodent and 2 in non-rodent in Japan. However, such descriptions were not existent in EEC and the USA (Speid et al. 1990). Because

statistically precise lethal doses are not necessary for regulatory purposes in Japan, we decided to delete the description from the new guideline (MHW 1993). Route of administration was also changed to focus on the clinical route.

Single-dose toxicity study is the only test by which we can observe the features of animals after high doses. By the precise observation, we can understand overall effects of the test substance and gain insight into the cause of death. These data were useful for (1) classification of the drugs for regulatory purposes, (2) determination of overdose risks following accidental ingestion of substances and the treatment at the time of intoxication, and (3) dose selection of repeated-dose toxicity studies.

With revision of the ICH M3 guidance (ICH M3 (R2) 2009), the need for acute toxicity testing was further refined and essentially eliminated except in very limited circumstances. It was concluded in extensive discussions that much of the essential information from these studies for pharmaceutical development can be derived from carefully monitored short-term repeated-dose toxicity studies or dose-ranging studies. This action further reflects the willingness for the ICH parties to harmonize to eliminate unnecessary animal use whenever feasible.

4.2.2 Toxicity Evaluation of Metabolites

When ICH was started, Japan was the only country that recommended the toxicity evaluation of the metabolites. At the first meeting of ICH in Brussels (1991), it was proposed by non-Japanese-based companies that this testing requirement be eliminated. However, there were many compounds that caused pharmacological and toxicological effects via metabolic conversion of the parent compounds, and significant species differences in the metabolic process were also well known. Therefore, the Japanese health authorities considered that toxicity studies on the major metabolites are, sometimes, necessary in the new drug application to make a reasonable extrapolation of animal data to human and, therefore, indicated in the first ICH meeting the circumstances under which toxicological evaluation of the metabolites was necessary (Ohno 1992). Those were major metabolites that may have toxicologically and/or pharmacologically significant effects and are specific to human, or its concentration in human blood at around clinical dose is much higher than that in animals used in toxicity studies. Criteria of the level of major metabolites were not mentioned. We considered that the decision should be made on case-by-case basis depending on the preceding information about the drug candidates.

Later in 2008, FDA issued Guidance for Industry "Safety Testing of Drug Metabolites." The guidance indicated that "Human metabolites that can raise a safety concern are those formed at greater than *10 % of parent drug systemic exposure at steady state*." However, there were many compounds like prodrugs for which exposure to parent drug is lower than those of the active metabolites. There was a possibility of request for quite minor metabolites. Therefore, this criterion was revised in ICH-M3(R2) guidance harmonized in 2009. It was described as "Nonclinical characterization of a human metabolite(s) is only warranted when that metabolite(s) is observed at exposures greater

than *10 % of total drug-related exposure* and at significantly greater levels in humans than the maximum exposure seen in the toxicity studies." It is noteworthy that in recent years, there have been several proposals for development of an ICH guidance on metabolite testing from industry and various regulatory authorities, in recognition of the increasing disparity of approaches and recommendations to this toxicity testing issue.

4.2.2.1 Repeated-Dose Tissue Distribution Studies (ICH-S3B)

Multiple-dose tissue distribution studies using radio-labeled compounds had been routinely conducted in Japan according to the guidelines on pharmacokinetics of pharmaceuticals (MHW 1991). On the other hand, those studies had been exceptional in the USA and EU. Therefore, ICH steering committee, held in Tokyo in September 1992, decided to start discussion aiming to harmonize the circumstance under which repeated-dose pharmacokinetics studies using radio-labeled compounds are necessary for drug development, and Japanese experts took the role of rapporteur.

The Japanese guideline did not necessarily require the use of radio-labeled compounds for the multiple-dose tissue distribution study; however, use of the radio-labeled compounds was essentially inevitable to do such a study. Therefore, it was not appropriate to indicate that the use of radio-labeled compounds was one of the choices of the researchers; it was essentially the only choice. On the other hand, upon re-examination of the 90 NDA documents submitted to Japanese authorities, it became apparent that repeated-dose tissue distribution study was not always conducted. The number of NDAs with a tissue distribution study was 72 and that with both single- and repeated-dose studies was only 47. That is, application of this guideline had been relatively flexible. It was noted that there were several compounds for which prediction of tissue distribution after repeated dose was not possible by single-dose studies (Table 4.2). When JPMA member companies were surveyed on the need for the repeated-dose tissue distribution study, only five pharmaceutical companies among 51 answered companies considered the repeated-dose tissue distribution studies unnecessary (Table 4.3). Based on these data and further discussion in ICH framework, a consensus was achieved on the circumstances under which the studies are necessary and ICH experts could reach consensus of ICH Step 4 in October 1994 (S3b: Pharmacokinetics: Guidance for repeated dose tissue distribution studies).

4.2.3 Purpose of Tissue Distribution Studies

Tissue distribution study aims to clarify the amount and nature of the chemicals related to the administered drug in tissues. These data are useful to the evaluation of organ responses from pharmacology and toxicology tests. The data are sometimes used for preparation of protocols for further pharmacology and toxicology testing. Before administering radio-labeled compounds for mass balance study in human, tissue distribution study is essential to estimate organ-specific exposure to radiation.

Table 4.2 Example of failure of prediction of tissue distribution after repeated dose by single-dose studies

	Single	7th	14th	21st
	(ng eq. of drug/g or ml)			
Plasma	12.5	28.9	31.7	26.8
Pituitary	N.D.	N.D.	N.D.	116.5
Thyroid	N.D.	160	245.8	457.3
Liver	299.9	956	1,208	1,000
Kidney	87.5	355.2	561.7	509

Concentration at 24 h after the last administration
Sano et al. Xenobiotic Metab. Disp. (1989)

Table 4.3 Survey on the necessity of repeated dose tissue distribution studies

Answer	Number of companies	Ratio (%)
Not necessary	5	9.8
Necessary	18	35.3
Necessary with conditions	28	54.9
Total	51	100.0

JPMA report 26th Oct 1992

On the other hand, purpose of the multiple-dose tissue distribution studies is to evaluate the steady state or accumulation after repeated administration. Therefore, pharmaceuticals used only by a single dose are not generally recommended to conduct repeated-dose tissue distribution studies. In addition, rough prediction of the results of multiple doses by single-dose study result is possible in most cases. Single-dose study is also enough where excretion is rapid and accumulations in tissues are not expected.

Repeated-dose tissue distribution studies had not always been required for all drug candidates. It should be considered depending on the nature and preceding information obtained. The ICH-S3B guidance was the result of intensive discussion and scientific debate of the six ICH parties.

4.2.4 Circumstances Under Which Repeated-Dose Tissue Distribution Studies Should Be Considered

The guidance indicated four circumstances under which the studies should be considered. The first was "when single dose tissue distribution studies suggest that the apparent half-life of the test compound (and/or metabolites) in organs or tissues significantly exceeds the apparent half life of the elimination phase in plasma and is also more than twice the dosing interval in the toxicity studies." This was because tissue concentrations are generally determined by plasma concentration, and we considered that the prediction of accumulation at steady state seemed to be erroneous in those circumstances and might become more than 3 times that of C_{max} after single dose. The second was

"when steady-state levels of a compound/metabolite in the circulation, determined in repeated dose pharmacokinetic or toxicokinetic studies, are markedly higher than those predicted from single dose kinetic studies." We considered that reliability of the prediction of tissue distribution after repeated dose by single-dose study was low in these situations and, therefore, it seemed necessary to confirm by animal experiments. The third was "when histopathological changes, critical for the safety evaluation of the test substances, are observed that would not be predicted from short term toxicity studies, single dose tissue distribution studies and pharmacological studies." We considered that safety evaluation in this situation should be conducted carefully by using information from various aspects. The repeated-dose studies conducted under these circumstances seemed to be critical for the safety evaluation of the pharmaceuticals, and therefore, the studies are recommended to be conducted under GLP regulation. The fourth was "when the pharmaceutical is being developed for site-specific targeted delivery." We considered that the characteristics of these pharmaceuticals should be confirmed by repeated-dose studies. Distribution to the other sites other than the targets should also be considered. Use of the same pharmaceutical formulation as that of the clinical trials was recommended.

In the case of high dose used in toxicity studies, the dose may be high enough to saturate the metabolic and/or excretion pathways, and the distribution after repeated dose may be deviated from the prediction by single-dose studies. The need for a repeated-dose tissue distribution study is not always supported in these circumstances and should be determined based on the toxicological significance of the study.

References

ICH M3(R2) (2009) Guidance on nonclinical safety studies for the conduct of human clinical trials of and marketing authorization for pharmaceuticals. 11 June 2009. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Multidisciplinary/M3_R2/Step4/M3_R2__Guideline.pdf

JPMA (1992) Survey on the relationship between clinical side effects and data from animal experiments. JPMA report No 61, 26 Oct 1992 (Original title was in Japanese and translated into English by Yasuo Ohno)

MHW (1989) Guideline for toxicity studies for pharmaceuticals. Yaku-Shin 1, Notification No 24 (11 Sept 1989)

MHW (1991) Guideline on non-clinical pharmacokinetics. Yaku-Shinyaku, Notification No. 6 (29 Jan 1991)

MHW (1993) Revision of single dose and repeated dose toxicity studies. Yaku-ShinYaku, Notification No 88 (10 Aug 1993)

Ohno Y (1992) Toxicity testing: Regulatory perspectives. In: D'Arcy PFD, Harron DWG (eds) Proceedings of the first international conference on harmonization, Brussels, Belgium, Nov 1991. The Queen's University of Belfast, Belfast, pp 186–188

Russel WMS, Burtch RL (1959) The principles of human experimental technique. Methuen, London

Sano H et al (1989) Structure on the fate of nizatidine (II) transfer into the fetus and milk, and absorption, distribution, metabolism and excretion after repeated administration to rats. Xenobiotic Metab Dispos 4:43–54

Speid LH et al (1990) Harmonization of guidelines for toxicity testing of pharmaceuticals by 1992. Regul Toxicol Pharmacol 12:179–211

Chapter 5
Toward More Scientific Relevance in Carcinogenicity Testing*

Jan Willem van der Laan, Joseph J. DeGeorge, Frank Sistare, and Jonathan Moggs

Abstract Carcinogenicity testing was chosen as one of the topics wherein harmonization could lead to more efficient guidance for the pharmaceutical industry without compromising human safety. An important difference in dose-selection strategy was the "toxicological" approach of the US FDA versus the "clinical dose margin" approach of the EU CPMP and the Japanese MHLW. The dose-selection guidance describes several acceptable approaches, including a new approach of the 25-fold AUC.

Discussion on the need for two species (rats and mice) led to the (initially unforeseen) introduction of transgenic mice as possible models of choice. In the intervening time since the guidances were first released, new developments in the carcinogenicity testing strategy are now seen as possible based on database evaluation and new insights in molecular biology of cancer mechanisms.

*The opinions expressed in this chapter are the responsibility of the authors only and do not represent the opinion of the organizations they are affiliated to.

J.W. van der Laan (✉)
Section on Pharmacology, Toxicology and Biotechnology, Medicines Evaluation Board, PO Box 8275, 3503RG, Utrecht, The Netherlands

Graadt van Roggenweg 500, 3531AH, Utrecht, The Netherlands
e-mail: jw.vd.laan@cbg-meb.nl

J.J. DeGeorge • F. Sistare
Drug Safety Testing, Merck Research Laboratories, West Point, PA, USA

J. Moggs
Discovery & Investigative Safety, Preclinical Safety, Novartis Institutes for Biomedical Research, Basel, Switzerland

J.W. van der Laan and J.J. DeGeorge (eds.), *Global Approach in Safety Testing*, AAPS Advances in the Pharmaceutical Sciences Series 5, DOI 10.1007/978-1-4614-5950-7_5,

5.1 Introduction

Despite several decades of research into understanding and developing interventions for prevention and treatment, cancer remains an important disease in the modern Western world with more than 1 in 4 lifetime risk for developing the disease. Further, with a few exceptions, there is still an increasing incidence in the types of various cancers (WHO and GLOBOCAN 2008), and while in many cases survival is improving, it has not yet become a "chronic disease condition" despite the development of a number of novel anticancer therapies (including anticancer pharmaceuticals). With a few exceptions (e.g., smoking, viral infection, alcohol consumption, and some chemical exposures), it is difficult to discern the causal agents. It is generally believed that at least some of the unattributed risk is as a result of environmental chemicals to which the human population is exposed essentially unavoidably. Of similar concern is that some of the risk may be posed by intentional exposure to chemicals as pharmaceuticals used in the treatment of various diseases. There is robust evidence that in some, limited cases, this concern is justified. The International Agency on Research on Cancer (IARC) has the task to evaluate the carcinogenic potential based on epidemiological and empirical (animal) data, and these datasets are important as the "gold standard" for reference for compounds for concern. This listing includes some pharmaceuticals that have been strongly linked to human cancer outcomes.

As a precautionary principle well established in regulation and industry practice, it is important to assess as early as practical the possible carcinogenic potential of the chemicals to which the population might be exposed. To address this, several general strategies have been implemented to avoid the unintended or unknowing introduction of chemical carcinogens into society use.

Since the 1960s, these preventive measures have included the requirement for testing new compounds in animals and evaluation of the outcome of these tests on cancer endpoints (WHO 1961, 1969). The protocols for testing for carcinogenic properties were developed in the middle of the last century with refinements following the introduction of Good Laboratory Practice. The current protocols which generally include lifetime testing at high doses in rats and mice are mainly based on the OECD Guidelines which came into force in 1979. There is little differentiation in testing method, regardless of the nature, application, or extent of understanding of the specific chemical of concern. This is a particular issue for pharmaceuticals where there is controlled exposure and specific patient benefit from use of the pharmaceutical and where there is extensive understanding of the pharmacology, general toxicology, and human experience generated during drug development that could clearly contribute in assessing potential carcinogenic risk.

At the start of the International Conference of Harmonisation, the topic of carcinogenicity testing was chosen as one of the topics wherein significant progress could be made by developing a unified guidance that factored in pharmaceutical-specific considerations.

At the first conference in Brussels (6–7 November 1991), an overview was given on the topic and several questions were developed around which revision of

the existing guidance was envisioned. Regarding the need for two species, there was already experience available at that time suggesting that a single-species test may be adequate for predicting human risk. The utility of the mouse bioassay in particular was highly criticized (Schach von Wittenau and Estes 1983), and this was explicitly expressed during this meeting (Hayashi 1992; Emerson 1992; Schou 1992).

It is important to note that by the early 1990s, there was already a substantial experience with the usual approach of lifetime studies in rats and mice in OECD Guideline 451, and several refinements have been proposed in the scientific literature and at workshops on carcinogenicity testing for pharmaceuticals. However, there was general agreement across the ICH regions that the then employed practice of lifetime testing was the most appropriate approach to test the carcinogenic potential of pharmaceuticals for human use.

This generally accepted approach can be summarized as testing any pharmaceutical with the potential for long-term use at a maximally tolerated dose in two species, usually rats and mice, but other species also being employed, for the anticipated life span of the animals. However, different regions had different views on details of the study design, for example, what constituted "long-term" human use and what doses to be used (especially for pharmaceuticals with a low toxic potential). ICH1 therefore recommended that a working guideline should be developed for rational selection of appropriate exposures and the corresponding doses (ICH 1992). It was stated that "the design of carcinogenicity studies, including the dose, number of species, and duration" needed to be reconsidered and "It is felt that there are fundamental questions about the rationale and criteria for current carcinogenicity studies which need to be examined" (ICH 1992).

In literature (e.g., IARC monographs), occasionally single species had been successfully employed to assess risk. Frequently, at the time of potential registration, only one of the two species was considered to have appropriately evaluated carcinogenic risk (Van Oosterhout et al. 1997; Contrera et al. 1997). Given this experience, to avoid unnecessary animal use in pharmaceutical testing, there was specific focus on reevaluating the utility of the routine practice of studies in two species for carcinogenicity assessment.

In total, these discussions yielded three work streams for developing pharmaceutical-specific carcinogenicity testing recommendations:

(A) Defining the conditions for a pharmaceutical that necessitated the specific conduct of carcinogenicity studies.
(B) Discussing the necessary constituents of the routine testing approach driven by an assessment of the value of the elements of the standard two-species lifetime design.
(C) Determining criteria for selection of doses that were more appropriate for pharmaceuticals in contrast to the maximum tolerated dose (MTD) used for general chemicals. Could pharmacodynamic or pharmacokinetic properties of specific pharmaceuticals, other than generalized toxicity, be used for dose selection and, if so, based on which specific considerations?

In the discussion that follows, the topics have not been addressed as in the order of guidances, but as in the order of ICH discussion prioritized them based on what could be most readily agreed. Dose selection was thus the first topic.

5.2 Development of a Guideline for Dose Selection for Carcinogenicity Studies ICHS1C

Rationale for a Dose-Selection Guidance: Carcinogenicity studies were and are amongst the most resource intensive and longest duration studies conducted as part of the nonclinical support for pharmaceutical development. It was recognized early on in the ICH process that establishing criteria in the design of these studies that would be universally accepted could eliminate a significant waste of animal and financial resources used in repeating studies to address different regional regulatory guidance. Both industry and the US FDA recognized that a substantial number of carcinogenicity studies were rejected by FDA as not being adequately designed. One of the most common causes for having a "failed study" was the failure to demonstrate that a maximum tolerated dose (MTD) or maximum feasible dose (MFD) was used in the carcinogenicity study. In many cases, this failure was the result of the industry aligning their practice with European or Japanese regulatory approach of accepting studies conducted at a ≥100-fold the clinical dose on a mg/kg comparative basis. This endpoint was not accepted by the US FDA, and instead studies conducted to these dose-selection criteria were retrospectively evaluated for having achieved either an MTD or MFD. Studies failing to achieve these later endpoints either needed to be repeated or other studies conducted to determine how close they had come to achieving an endpoint acceptable to the FDA. This divergence of regulatory posture was thus driving industry behavior and resulting in not infrequent additional expenditure of resources. Thus, the fundamental premise for creating guidance for high dose selection in carcinogenicity studies was rationalization of dose-selection criteria across the ICH regulatory regions with clear delineation of uniformly acceptable criteria.

5.2.1 Issues in Achieving a Unified Dose-Selection Guidance

On the surface, achieving a unified ICH guidance could have been as simple as agreeing on mutual recognition of the existing dose-selection criteria by all regions. However, at the first meeting of the ICH in November 1991, it was declared that "neither an MTD nor an arbitrary multiple of the clinical dose" should be used to select the high dose for carcinogenicity studies (ICH 1992). The mandate was to develop a science-based rational approach specifically relevant to human pharmaceuticals. Therefore, the mutual recognition of existing approaches was not an

option. An equally important hurdle was that distinct nonclinical regulatory philosophies appeared to exist in the different authorities underlying these dose-selection criteria.

At the US FDA, it was felt that toxicity studies were to evaluate the full range of the toxic potential of a compound, regardless of relevance to clinical use. Once this profile was fully evaluated, the interpretation of relevancy of findings to human risk could be considered, but this was secondary to observing the full spectrum of chemical toxicity. Thus, FDA has felt compelled to conclude that even in the presence of some tumor findings in rodents a positive risk benefit analysis often resulted, even for nonlife-threatening disease. However, FDA almost never declared that the tumor observations were irrelevant for humans. Regardless of the test conditions under which the rodent tumors were observed, in nearly every case, the findings were listed in the product label.

European and Japanese regulatory philosophy had as a major focus identification of those risks primarily in a range of doses considered directly relevant to clinical use of the pharmaceutical. Provided there was a significant "margin for findings" to clinical use, the observations could be considered of minimal clinical concern and in fact need not be identified. This margin approach focused on dose, a practice common at the time, not on systemic exposure as currently used. The ICH guidance on toxicokinetics (ICH 1994) had not been crafted yet, and there was usually only minimal information collected on systemic exposures achieved in toxicity studies. Thus, the emphasis was on limiting the high dose used in toxicity studies in relation to the clinical dose, and there was a general lack of concern for toxicity or tumors that might occur above the declared arbitrary dose margin.

Uniformly, the industry's position on evaluation of toxicity was more aligned with that of European and Japanese regulatory authorities. In the case of carcinogenic potential and toxicity testing in general, it was an industry preference to investigate findings and access risks specifically at doses within the pharmacodynamic range of the test species. Since at the time almost all human pharmaceutical targets had relevant animal models that could be relied upon to determine appropriate, pharmacodynamically active doses in rodents, this was not then the issue that it would be today. The industry view (and to some extent the European and Japanese regulatory view) of limiting doses to the pharmacodynamic range was driven by the belief that effects observed beyond the pharmacodynamic range were off target, should be unattained in clinical use, and thus irrelevant to the patient's risk. This opinion is elegantly elaborated upon by Monro (Monro and Mordenti 1995), one of the principal industry ICH S1B EWG members for the S1C guidance. The industry representatives were generally aligned on elimination of MTD, MFD, and high arbitrary multiples of the clinical dose as criteria for high dose selection. In fact, the MTD endpoint was considered by the industry a disadvantage, or in some cases a penalty, in developing drugs that were of low toxicity in rodents compared to those which were significantly toxic at small dose margins to the human therapeutic dose. A similar view also played a part in the rationale for the EU and MHLW support of a high dose multiple (100×) as a dose-selection endpoint for carcinogenicity studies.

5.2.2 *Bridging to a Uniformly Acceptable Guidance*

Given the philosophical differences in starting positions, progress was initially slow and yielded little success. Progress was initiated when Contrera and colleagues (1995) conducted an analysis of the exposure and dose used in rat carcinogenicity studies conducted at the MTD and compared them to clinical exposures at the human pharmacodynamically active or efficacious dose for a number of pharmaceuticals that the FDA had reviewed. While the dataset was relatively limited, the surprising result was that at the MTD exposures were not routinely excessively high compared to the clinical therapeutic exposures. Approximately 1/3 of compounds yielded exposure multiples in rodents of the human exposure of 1 or less, of between 1 and 10, and of greater than 10, with few compounds producing exposures multiples of greater than 50-fold that of the clinical exposure. An additional important observation from the analysis was that the pharmacokinetic systemic exposure multiple achieved was approximately predicted when the dose data were normalized and compared on a mg/m^2 dosing basis. This latter insight allowed extension to a substantially larger sample of pharmaceuticals for which pharmacokinetic data were not available from rodents and confirmed the distribution of estimated exposure multiples achieved in carcinogenicity studies for which pharmacokinetic data were available.

Overall, this analysis helped to change the mind-set of EWG members in several ways. Most importantly, it was not feasible to eliminate the MTD as an endpoint, as many compounds could not be delivered to achieve substantially greater exposures in rat than were achieved in patients. This assumes that the relevant tissue compartment's exposure is reflected by the systemic plasma compartment exposure. As noted above, 2/3 of the compounds tested at the MTD achieved exposure multiples of tenfold or less of the clinical exposure. None of the EWG parties considered this exposure multiple excessive. (Some participants still considered the effects generated at the MTD as a distortion of the properties of the drug under pharmacologically irrelevant conditions, but had no viable alternative recommendation.) Differences in the philosophies between the industry's desire to focus on pharmacodynamics, EU and Japanese regulators on safety margins, and the FDA regulators on the full profile of toxicity became irrelevant for a substantial proportion of pharmaceuticals, as regardless of the philosophy the maximum dose that could be tested apparently yielded exposures within what was a generally acceptable range for all parties. This resulted in a modification of the discussion of the MTD from how to eliminate it as an endpoint to a focus on developing a practical, harmonized prospective definition. The analysis also made it apparent that the EU and Japanese approach of allowing the high dose to be defined as 100-fold the clinical dose on a mg/kg basis was projecting a 15- to 20-fold exposure margin either by pharmacokinetics or based on dose normalization to mg/m^2. This realization along with another observation from the Contrera et al. (1995) analysis indicating that rodent testing identified clinically relevant carcinogenic risks at exposure multiples within a 20-fold the clinical exposure helped support a potential dose-selection endpoint

of a 20- to 50-fold exposure to the clinical exposure. Other than the units employed and the scientific underpinnings, this did not yield substantially different upper dose selection from that of the European and Japanese authorities' traditional dose-fold approach.

Based on the analysis, a general principle that could be agreed on was that "ideally the doses selected … should provide an exposure to the agent that" yielded an adequate safety margin relative to the human exposure, was tolerated without chronic physiological dysfunction, focused broadly on the properties of the agent in human and rodent, and enabled interpretation of results in the context of human use. With this general agreement came the realization that no one dose-selection approach was likely to address all of these aspects for all compounds and equally that no one dose within a study would provide the necessary context to interpret a study's relevance. The outcome of this shared understanding was that high dose-selection criteria would need to be flexible and advice on how to set the mid and low doses, not initially part of the EWG's work plan, was necessary. Although not all EWG members, particularly some industry representatives who wanted more focus on pharmacodynamics, agreed with all the conclusions being drawn from the work of Contrera et al. (1995), it opened the door to a new dialog that became the foundation of the guidance.

5.2.3 High Dose Selection

The step 2 draft version of the guidance released for comment (Fed. Reg. 59, 1994) specified four alternative approaches to high dose selection: pharmacodynamic endpoints, toxicity-based endpoints, pharmacokinetic endpoints, and saturation of absorption, as well as a statement to consider additional, nonspecified endpoints on a case-by-case basis. For the latter, with the exception of mentioning a C_{max} alternative pharmacokinetic endpoint and other nonspecified toxicity endpoints, there was no guidance on what these other endpoints might be, except to state that other endpoints not yet known may have merit and would need specific justification.

5.2.3.1 Pharmacodynamic-Based Endpoint

For the standard endpoints proposed, in an attempt to de-emphasize the use of the MTD, the pharmacodynamic endpoints were discussed first in the document. The potential pharmacodynamic endpoints were considered to be highly variable, compound specific, and dependent on the pharmacological selectivity of a given compound. The definition of what an appropriate pharmacodynamically selected high dose might mean, however, suggested a significantly limited application that was linked to pharmacologically driven toxicity. It was to be a dose "not producing disturbances in physiology or homeostasis … but should produce a pharmacodynamic response … which would preclude further dose escalation…" This definition was

viewed by some EWG members as little more than a pharmacological target-based MTD and not necessarily addressing the intent of the industry proposal for a pharmacodynamically driven high dose-selection consideration. The definition was largely unchanged in the final version of the guidance but had examples added to the text that make it clear that these are in essence "toxicity" limitations on increasing the dose driven by significantly adverse pharmacology. In recognition of this minimized role and close relation to standardly accepted toxicity, the pharmacodynamic endpoint was moved to the second to last endpoint discussed in the final guidance. The toxicity-based MTD, in contrast, was discussed first in the final guidance in recognition of its likely primary application in dose selection. This could be viewed as a failure of the guidance to achieve the initially stated objective but in fact was more a recognition of the impracticality of those initial objectives.

5.2.3.2 Toxicity-Based Endpoint: MTD Discussion

While the work by Contrera et al. (1995) made it clear that an MTD would need to be maintained as an option, it did not contribute to determining which definition of the MTD would be used. Captured in Note 1 in both the draft and final S1C guidance are several of the existent definitions of the MTD available at that time from various government and regulatory groups. In sum, the definitions of the MTD in some aspects appear conflicting (e.g., "causes no more than a 10% decrement in weight gain" vs. "should produce a 10% weight loss or failure of growth"). In others aspects, the MTD seems to be identifiable only in retrospective examination of the completed bioassay study in having been exceeded. While this was useful in evaluation of a study, it was less valuable in prospectively designing a study that would use acceptable doses and be considered a valid study. This later point was of considerable concern, as it had caused a routine practice in industry of overshooting the MTD to clearly demonstrate that it had, in fact, been achieved. Originally, the EWG did not provide a definition of the MTD as is apparent in the published draft guidance, but instead stated that all of the referenced definitions provided as Note 1 were equivalent and thus equally valid (Fed Reg. 59, 1994). Even the term "MTD" was an acronym derived from different words with similar but not identical intent in the different regions. In the EU, MTD meant "minimally toxic dose," whereas in the USA, it meant "maximally tolerated dose." The comments to the published draft, however, indicated that the definitions available were unclear and contradictory (as noted above) and that calling them "equivalent" did not improve the utility of the MTD endpoint. To address these comments, the EWG crafted its own definition of the MTD that made it clear that a dose chosen as the MTD was to be evaluated prospectively, that is, "was a top dose … which is **predicted** to produce a minimally toxic effect over the course of the carcinogenicity study" (emphasis added). It further created a clear minimum definition of what constituted an appropriate prospectively selected dose and provided additional flexibility for using specific toxicity endpoints not generally incorporated in the previously stated definitions. There was still an attempt by the EWG, however, to not contravene the previously existing

MTD definitions, and it was stated in the guidance that the definition provided was still "considered consistent with those published previously by international regulatory authorities." In retrospect, inclusion of this statement has continued to cause confusion, implying that the other definitions are interchangeable with the ICH definition, which they clearly are not.

5.2.3.3 Pharmacokinetic Endpoints

The most novel and useful dose-selection criteria created in the ICH S1C guidance are the pharmacokinetic-based high dose endpoints, the 25-fold multiple of the clinical exposure, and the saturation of systemic exposure (see later). While it can be considered that the 25-fold multiple is a derivative of the 100-fold the clinical dose approach previously used in the EU and Japan, no similar exposure endpoint had existed in carcinogenicity dose selection, and none does outside of application to pharmaceuticals.

The development of the pharmacokinetic endpoint as a multiple of the human exposure was enabled by multiple considerations, analysis of numerous datasets, and significant compromises among the ICH parties to reach agreement. One of the first and most critical compromises was the acceptance that plasma systemic exposure calculated as the free drug area under the curve (AUC) would be the basis for the pharmacokinetic endpoints. This was a compromise, as comparisons of systemic exposure across species could not be clearly demonstrated to predict equivalent carcinogenic risk nor could the plasma compartment free drug concentrations be definitively demonstrated to best represent the variety of tissue compartments of free drug concentrations which would result in the carcinogenic risk. It was, however, considered the most reasonable assessment of comparative body burden and was considered to reasonably correlate with the types of nongenotoxic carcinogenicity mechanisms that could come into play in pharmaceutical-based carcinogenicity (e.g., immunosuppression, hormonal effects, and repeated organ insult). Once this was agreed, the next major hurdle was establishing what fold of exposure would be appropriate. The dataset analyzed for this purpose and criteria applied are presented in Note 4 of the Step 2 draft guidance (Fed. Reg. 59, 1994). The first criteria, "an adequate safety margin," is in part related to the European and Japanese approaches of 100-fold the dose on a mg/kg dose basis. When normalized to mg/m^2 dose comparison, an approximation that was used to normalize the comparative exposures across species in assessing the pharmaceutical carcinogenicity database (Contrera et al. 1995), this converts the approximate 100-fold dose ratio to an 18- to 20-fold and 8- to 10-fold estimated systemic exposure ratio for rat and mouse, respectively. Thus, acceptance of 25-fold multiple can be considered to retrospectively "validate" the adequacy of the previously used 100-fold the clinical dose.

There was a substantial discussion about alternatively accepting a 10×–15× exposure ratio. This discussion focused on two countervailing views. The historically accepted 100-fold of the clinical dose endpoint as having provided an adequately protective margin in the past yielding an exposure margin in this range

versus a concern that a margin of 10× for a carcinogenic risk was not adequately protective of human health. Criteria were then agreed by the working group that the margin would need to enable detection of known and probably human carcinogens, and this would establish a lower bound for an acceptable margin ratio. This group of known and suspected pharmaceutical carcinogens were mostly constituted of genotoxic compounds; one pharmaceutical for which there was adequate exposure information, phenacetin, appeared to need an exposure multiple of 15-fold the clinical exposure to be detected as a carcinogen in the rodent bioassay. The remaining pharmaceuticals from this group, most of which did not have adequate systemic exposure data, could be calculated based on a mg/m^2 normalization to have been detected as carcinogenic using multiples of <20-fold the clinical exposure. The discussion became one of how much of an additional "safety factor" should be applied, but this data essentially put a floor at 20-fold multiple. In light of this, it was proposed that a 50-fold margin be used; however, it was in the end agreed that the 25-fold margin would be sufficient as was proposed.

The dialog as to what the exposure margin should be continued after the publication of the draft guidance. Upon reevaluation of the data by PhRMA and FDA EWG members wherein the lowest dose producing a notable tumor response was evaluated (rather than assessment of the top dose used in the study), it was determined that application of a 10-fold exposure margin would have identified all the carcinogens with the exception of phenacetin which still required an exposure multiple of 15-fold. Despite this reanalysis, the 25-fold margin was preserved in order to ensure that an adequate safety factor existed for this new approach. The phenacetin multiple needed was further questioned and recalculated by the Swedish MPA colleagues (Bergman et al. 1998) by conducting new pharmacokinetic studies in rat. They concluded that the doses of phenacetin used previously yielded an exposure ratio of 7. The relevance of this data to the original study could be questioned, and given the limited impact of revision to the recommended ratio in the guidance, it was considered to minimal to justify guidance revision.

5.2.3.4 Pharmacokinetic Endpoint: Saturation of Exposure

The saturation of exposure endpoint is in the view of some a pragmatic but more rigorous application of the maximal feasible dose. This endpoint was immediately considered useful and of limited controversy. Once it was agreed that AUC would be considered the most practical way to measure "internal" dose, it made no sense to any of the EWG participants to continue to escalate to higher doses when internal exposure had ceased to increase. While discussed at the time, the EWG did not define "ceasing to increase the exposure with increased dose," which in practice is asymptotically achieved with increased dose. There was also no guidance offered on the efforts needed to demonstrate that altering formulation or dosing regimen would not further increase exposure. This lack of guidance has recently been partially addressed in the question and answer for ICH M3 (R2) Guidance (ICH Web site, June 2009) as an effort to improve guidance implementation in relation to using

the maximal feasible dose. In discussing the effort to demonstrate a "maximum feasible dose," the Q&A indicates that the intent is actually to maximize exposure and, thus, the answer is equally applicable to the saturation of systemic exposure endpoint. Other than this inferred guidance, there is no recommendation on what constitutes a convincing argument for demonstration of achieving the saturation of exposure endpoint.

5.2.3.5 Other Endpoints Considered

While repeatedly discussed, there was intentional omission of the percentage of drug in diet as a dose-selection endpoint. This has been used routinely as an endpoint for food and environmental safety testing and, historically for pharmaceuticals, as an upper bound dose based on concern for an impact on animal health. This dietary consumption endpoint was considered inappropriate criteria for a human pharmaceutical, as opposed to an environmental chemical or food additive, due to the nature and intent of pharmaceutical use, and was rejected as an endpoint worthy of inclusion in the new ICH dose-selection guidance.

5.2.4 Application of Metabolism Data in Carcinogenicity Dose Selection

Once it was agreed that an exposure multiple was an appropriate endpoint, the question became exposure multiple of what? Differences in the extent of metabolism between test species and humans have been widely recognized since metabolite profiling had been undertaken in the late 1980s as part of drug development. In cases where the vast majority of the systemic exposure in humans and the test species was to the parent drug, there was no question in how to calculate the margin. Use of the parent drug exposure alone was acceptable. However, when metabolites were significantly formed and circulating, the approach to calculating an acceptable margin was less clear. Three alternative positions were put forward by various members of the EWG (1) Only the parent compound should be considered as it was still the primary active agent. (2) The parent and all significant drug-related compounds should have a summated AUC and be considered as a whole in the calculation. (3) Each drug-related compound should be considered independently and each should achieve the proposed exposure margin. This last proposal was recognized as the least achievable and inevitably would have allowed very few, if any, compounds to be tested using the exposure-based endpoint. The first was the simplest and was the basis of deriving the 25-fold margin in the first place, as metabolites were not considered in the calculations of Contrera, except as approximated when using the mg/m^2 normalization. However, when faced with knowledge that significant differences in metabolism across species did exist, ignoring these differences could not be scientifically justified. In the end, the aggregate AUC approach was accepted.

In most cases, the exposure multiple was driven primarily by the parent drug, simplifying the calculation to a calculation of parent-only exposure. In those cases where extensive differences in metabolism across species were evident and where they contributed substantially to the overall exposure, the inclusion of metabolites in the assessment was considered valuable. This agreement maintained the utility of the exposure-based endpoint as one with broad application.

As noted above, comparative metabolism data was an important consideration in the development of the S1C guidance. As a general recommendation, it was agreed that species (or strains) selected for use in carcinogenicity studies should generate similar drug metabolite profiles to that observed in human. This concept, which on its face seems obvious, presented controversy within the EWG. A primary concern was that if none of the rodent strains evaluated had a "similar" metabolite profile with human, practically, there was relatively little that could be done. The species available to test with adequate historical carcinogenicity testing experience were relatively limited. Thus, the likelihood of identifying a strain-specific drug metabolite profile comparable with human was considered low if more traditional strains did not generate the necessary similarity. It should be noted that it was not a contemplated remedy by the EWG that separate carcinogenicity studies would need to be conducted with a "unique" or a "disproportionate" drug metabolite alone, as has recently been suggested and undertaken based on some regional health authority guidance (FDA 2008). Rather, the EWG considered this to have pragmatic solutions and this serves as the basis for this (and other) recommendations in the guidance being qualified by terms such as "ideally" or "as possible." It was clear to the EWG members that it would not always be possible or feasible to apply the recommended criteria and that this could still lead to an acceptably conducted study, provided the interpretation of the study outcomes considered these less than ideal circumstances. Unfortunately, it does not appear that this intended flexibility in study conduct is today still fully appreciated. Often, the recommendations in the guidance are relatively rigidly interpreted and adhered to by various regulatory authorities. The flexibility in metabolite comparability overall played a lesser role when determining if the exposure multiple approach was acceptable. As mentioned above, it was felt that there should be an assessment of comparable metabolite exposure, preferably in vivo, but at least as demonstrated by in vitro data. In the absence of comparable metabolite generation, the use of the exposure-based endpoint was not generally considered acceptable.

Another relatively new concept in this guidance is consideration of protein binding when assessing comparative exposure, whether applying the pharmacokinetic endpoint or not. As noted earlier, the use of exposure (and specifically the unbound plasma compartment exposure) as a surrogate for assessing carcinogenic risk was controversial within the EWG, even in the final guidance. This can be understood from the qualifications included in acceptance of the pharmacokinetic endpoint "the unbound drug is *thought to be* the most relevant," "no validated scientific basis for use of comparative drug plasma concentrations," and "is considered pragmatic." Inclusion of such language in the guidance highlights the divergent opinions, but did not prevent the relatively strong recommendations that underpinned exposure

assessments. Despite this stated agreement to use unbound fraction for comparison of exposure, Note 9 of the guidance makes it clear that this primarily applies when using the unbound fraction in calculations provided such consideration decreases the margin. Thus, the statement that using the total exposure "is acceptable if the unbound fraction is higher in rodent," but the note indicates "the unbound fraction *should be used*" when the unbound fraction is greater in human. There is no explicit acknowledgement that the margin ratio can be (or should be) calculated from unbound fraction when the rodent unbound fraction is greater. This has left this an open question, which in practice appears rarely accepted by regulatory authorities, amplifying the lack of conviction in application of the unbound fraction, unless it delivers a more conservative risk assessment.

5.2.5 Lower Dose-Selection Advice

It was recognized that the high dose selection was critical in elucidating the carcinogenic potential of the pharmaceutical. Whether the high dose selected was based on MTD, pharmacodynamics, or pharmacokinetics, it was unlikely that it could simultaneously provide complete information on the relevancy of any tumors observed for clinical use. For this evaluation, the middle and low doses use in the carcinogenicity study needed to be carefully selected to fully understand the response range and association with pharmacodynamics, pharmacokinetics, or toxicity. Traditionally, the middle and low doses used in carcinogenicity studies were fixed fractions of the high dose (a progression of 1/2 to 1/3 from high to middle and middle to low dose). For pharmaceuticals, to aid in understanding the interplay of nonlinear systemic exposure, development of off-target pharmacodynamics, and impact of organ selective nonlethal toxicity to the carcinogenic response and human risk, it was felt by some EWG members that the use of arbitrary multiples of the high dose should not be employed. While a proscription against the use of arbitrary multiples was not incorporated into either the initial draft or final guidance, an admonition to consider a broad range of criteria was incorporated. Unfortunately, this has not been sufficient to change either the behavior of regulators or the industry, and it is still routinely observed that "uniform dose spread" rather than mechanistic understanding drives the selection of the middle and low doses.

5.2.6 Modifications of the Guidance

5.2.6.1 Addition of Limit Dose Definition

In the final version of the original S1C guidance, there was discussion of as yet undefined dose-selection endpoints that should be justified on a case-by-case basis. Unlike the Step 2 version, however, there were no examples of what these endpoints

might be. Instead Note 11 in the final guidance made reference to an ongoing dialog for pharmaceutical-specific endpoints still in discussion. No such endpoints have been brought forward in the nearly 20 years since this statement was made, with one possible exception, the limit dose. The limit dose was proposed as an absolute cap on the dose to be tested in the rodent carcinogenicity study. While it had been general practice to limit non-pharmaceutical carcinogenicity testing to doses of 5,000 mg/kg as a component of diet in consideration of the impact on nutrition, a similar dose limit had been intentionally excluded in the S1C guidance. This had left as a case-by-case determination what dose could be used as an absolute maximum when none of the other defined acceptable endpoints had been realized. As indicated in the note on the S1C(R) guidance revision, however, this had been a very rare circumstance even without the flexibility that the guidance now offered. The industry had proposed a limit dose of 1,000 mg/kg, which was consistent with other toxicity testing guidance (ICH S5A 2000). The analysis of the FDA database of over 900 pharmaceuticals indicated that only 20 compounds had been tested at doses of 1,000 mg/kg or greater, with 7 of these positive only at or above the 1,000 mg/kg dose. The data analysis indicated that using doses of a maximum of 1,500 mg/kg would detect all carcinogens of concern. A further caveat on this limit dose endpoint was that it only applied to pharmaceuticals dosed in humans at 500 mg/day or less and indicates that the maximum feasible dose be used for drugs dosed at higher 500 mg in humans. As described in the Note 2 of the revision, this 500 mg maximum human dose was justified based on the mg/m^2 normalization between humans and rodents and a desire to maintain the 25-fold multiple when the 1,500 mg/kg dose in rodent is used. This endpoint took nearly 2 years to finalize and has been only infrequently used, but does provide an upper limit calculation for drug supply needs for carcinogenicity studies, and thus can facilitate planning during early development.

5.2.6.2 Removal of the Restriction for Using the 25-Fold Margin to Nongenotoxic Compounds

Recently (2008), S1C was again revised as S1C(R2). The primary revision was the removal of the restriction for using the exposure multiple endpoint only of pharmaceuticals without a genotoxicity signal. On the face of it, this revision can be questioned as to why a 25-fold exposure multiple without evidence of carcinogenicity is adequate for drug that has been shown to pose a genotoxic risk. Is a 25-fold margin with an absence of evidence of carcinogenic risk truly adequately protective of human health? For the answer to this question, one needs only to look at the original basis for the proposed 25-fold margin. The datasets of compounds were those which were known or suspected human carcinogens (e.g., also including phenacetin) and for which the 25-fold margin was considered adequate for detection. These compounds with known or suspected risk were primary genotoxic carcinogens. Thus, the original exclusion of genotoxic compounds from this testing endpoint was not scientifically justified, and the revision rectified this original oversight. While there

were numerous other minor changes in the S1C(R2) version of the guidance, most were either legalistic changes, "may" to "can" edits, or deletions of text relevant to the deletion of the genotoxicity restriction. The revision did not take the opportunity to correct any other deficiencies in the guidance.

5.2.7 Opportunities

5.2.7.1 Dose Selection for Transgenic Mouse Models

A primary failure of the guidance was a failure to include any discussion of dose selection for carcinogenicity studies in transgenic mice. The acceptability of the intermediate duration transgenic mouse as a test model instead of the 2-year mouse bioassay (S1B) was not completed until after the implementation of the S1C guidance. Thus inclusion of transgenic animal dose selection could not even be contemplated at the time of the original guidance. However, revisions of S1C that were occurring either simultaneously with or several years after S1B guidance that allowed use of transgenic animals were finalized. No mention of what endpoints could be acceptable for transgenic mouse studies is available in either guidance. In practice, the only endpoint accepted by regulatory authorities is the MTD. This has significantly limited the utility of the transgenic mouse as an alternative model for the same reasons alternative dose-selection endpoints for 2-year bioassays have been improved by the availability of alternative dose-selection criteria for 2-year bioassays.

5.2.7.2 New Developments

The original Note 11 (now Note 12) speaks of active discussion of alternative pharmaceutical-specific endpoints. With much recent focus on pharmacodynamics as providing insight into relevant carcinogenic risk, and the application of toxicogenomics as potentially contributing to cancer risk identification and assessment, there is no ongoing dialog as to how these may be factored into dose levels and more general design issues for these studies. The innovation in the toxicological assessment of pharmaceuticals initiated in the 1990s has essentially stalled in the early 2010s.

5.2.8 Value and Impact of the S1C Guidance

The original intent of the ICH1 conference and declaration that carcinogenicity study design and assessment needed revision to make it more useful and minimize resource wasting, especially animal use, was noble. The focus on dose selection

as an opportunity to generate harmonized study designs that would reduce the occurrence of unnecessarily repeating studies was laudable. The stated proposal to eliminate the use of the "MTD or an arbitrary multiple of the clinical dose," however, was misguided. The S1C guidance established the acceptability of a pharmaceutically relevant MTD, created, if not an arbitrary, at least practical and experience-based multiple of the human clinical (dose/exposure) as a carcinogenicity endpoint, and created the flexibility to use other practical endpoints for selection of the dose range used in carcinogenicity studies. In sum, there were a number of reasonable, data-based assumptions made in the development of exposure and other criteria as endpoints for carcinogenicity study dose selection. These assumptions could only be tested in a limited manner, and yet they were important in underpinning the guidance. It was for this reason that this specific exposure-based endpoint as defined in the guidance was considered and stated as "pragmatic," but a similar pragmatism ran throughout the guidance, even while it broke new ground in regulatory recommendations of carcinogenicity testing.

In terms of value, the guidance created a framework for dose selection for the most resource intensive studies conducted in the nonclinical development of pharmaceuticals that radically limited the repeating of studies based on "inadequate doses being used." FDA which had rejected numerous carcinogenicity studies prior to the guidance as having inadequate dosing has in the years since rejected none when using the defined endpoints and prospective consultation on the dose levels (personal communication). Moreover, experience has demonstrated that careful application of the dose-selection criteria (including having FDA independently validate the criteria) can generally assure global acceptance of a study conducted using the criteria. While this guidance clearly could be further improved (which has been pointed out throughout the proceeding discussion), this guidance has delivered on its intended objectives.

5.3 S1A Need for Carcinogenicity Studies

While there appeared to be general agreement on which products needed an assessment of carcinogenic potential, there was enough divergence that an EWG discussion was considered necessary to define for which circumstances a full carcinogenicity study package would be warranted. There was agreement on the main criteria, but some details were insufficiently spelled out. The main issues were:

1. Cause for concern, for example, compounds with genotoxic features, evidence of preneoplastic toxicity in repeated dose toxicity studies
2. Duration of the clinical therapy and thus duration of exposure of the patients

Other aspects considered were indication and patient population (e.g., compounds for a life-threatening disease) and route and extent of systemic exposure necessary when the clinical route was other than the oral route. In general, it was felt that these aspects were less controversial and played a minor role in the discussions around carcinogenicity studies.

The last issue was whether carcinogenicity studies would be needed for endogenous peptides and other protein substances. This issue was taken on board by the ICH S6 expert working group, the first ICH guideline being released in 1997. The outcome was that in general, carcinogenicity studies do not have additional value in view of the known pharmacological properties of these compounds. See further discussion on the S6 guidance in this book.

5.3.1 Cause for Concern

Parallel to the discussions in the EWG on carcinogenicity, there was also an EWG on genotoxicity. This genotoxicity group was establishing a standard battery of tests to define the genotoxic character of human pharmaceuticals.

There was and remains a consensus that the main outcome of genotoxicity is the induction of DNA damage in the somatic cells and that genotoxicity enhances the carcinogenic risk more consequentially rather teratogenic, reproductive risk. Most genotoxic compounds (approximately 90%) induce tumors after long-term use, although this leaves 10% of genotoxic compounds as exceptions. In line with this observation, most of the IARC class 1 and 2A compounds are genotoxic.

Recognizing this, the EWG proposed that evidence of significant genotoxicity (as established after evaluation of the compound in the standard battery, sometimes with extended testing) can be taken as sufficient information to decide that there is a significant carcinogenic risk. Long-term testing in two species was decided as inappropriate in such cases, as in most instances it would only confirm the well-understood risk of the compound. It was already expressed several times in the ICH process that if the outcome of a study is largely predictable, such a study would be pointless (Monro 1994). The conduct of a bioassay with a highly predictable outcome is difficult to defend on the basis of not generating new scientific information, the unnecessary use of animals, and the resources expended.

Significant genotoxicity is a cancer risk. What is the value in demonstration of this in long-term studies? This recommendation is important in that it helped reduce unnecessary studies. The conclusion that evidence of genotoxicity is primarily a cancer risk rather than a reproductive risk was confirmed recently in the ICH M7 discussions, where it is agreed that the discussions on genotoxic impurities are important primarily in relation to cancer risk.

In the area of non-pharmaceutical compounds, it is common to calculate the potency of a compound. Hernandez et al. (2011) have calculated a quantitative relation to predict carcinogenicity from evidence of genotoxicity in vivo. Although there are limitations to this approach, because of the small number of studies, they have described a strong correlation between the potency to induce DNA damage and the resulting carcinogenicity. These data and the analysis confirm the approach chosen by the EWG almost 20 years earlier.

The wording of the S1A guidance includes also “evidence of preneoplastic toxicity in repeated dose toxicity studies” as a cause of concern. While the first mentioned cause for concern, genotoxicity, might result in not doing a study (because

the carcinogenic risk is anticipated), evidence of preneoplastic toxicity is taken as an indication that this should be a reason to conduct a full carcinogenicity assay, in order to assess the potential progression to cancer illustrated by the preneoplastic findings. This recommendation is important in view of the recent discussions about the predictability of carcinogenicity testing outcome based on pharmacological and toxicological properties of the compound (including absence of evidence of preneoplastic lesions). We will discuss this again at the end of the chapter.

5.3.2 Duration of Clinical Therapy

For nongenotoxic compounds, duration of treatment (long-term exposure) is thought to be important to the level of carcinogenic risk posed. Different standards for the duration triggering testing were imposed in the different regulatory regions, but the scientific basis for the difference between 2–3 months (FDA) and 6 months (the EU and Japan) was unclear. What appeared to be initially an intractable difference to bridge was solved easily, but not directly by the toxicological experts, but rather by clinical practice. In clinical practice, there is an obvious differentiation between long-term and short-term treatment. Short-term treatment might be a single administration (as with diagnostics) or just with a week or month duration (as with antibiotics), but treatment schedules with a longer duration are also likely to be repeated, adding up to a likely duration in the order of magnitude of several months within a few years, and may be more over a lifetime, suggesting a risk commensurate with that of repeated long-term administration.

From a scientific point of view, interruption of treatment may lead to reversal of the effects and decreased proliferative responses. This would be contrary to the assumption that repeated intermittent administration of a compound would lead to an accumulated risk for proliferation and carcinogenicity. However, other theories support the concept of accumulation of risks after intermittent exposure. In the absence of specific evidence for any given pharmaceutical and its mechanism of nongenotoxic carcinogenesis, the S1A guidance took a conservative approach, covering the possibility of an accumulating risk. Thus, the guidance makes the recommendation that pharmaceuticals for use in repeated short-term treatment of chronic recurrent disease, such as antihistamines for seasonal allergy, should undergo testing similar to those pharmaceuticals for chronic continuous treatment.

5.4 S1B Two Species

5.4.1 Background of Choice of Two Species

A first global agreement on testing on carcinogenicity was reached within the framework of the WHO as early as in 1961. In a technical report (WHO 1961), recommendations have been given regarding numerous details of carcinogenicity studies of food substances. From this report is the following statement:

> Both sexes of each of at least two species of animals should be used in the tests throughout their life span. In most cases these species would be rats and mice. Hamsters or dogs might be suitable, but guinea-pigs, for example, appear to be resistant to some known carcinogens. The use of dogs in carcinogenicity tests has disadvantages. Because of the expense of maintenance it is difficult to use a sufficient number to detect the low incidence of cancer, and the life span of this animal is 12–15 years.

It was therefore pragmatic that testing for carcinogenic potential would be conducted in different species, but for practical reasons just two rodent species are the standard and not a rodent and a non-rodent (as for repeated dose toxicity or reproductive toxicity).

This choice of two species was confirmed in a Technical Report on Carcinogenicity Testing of Drugs (WHO 1969).

However, the value of the mouse was already disputed as early in 1972 (Grasso and Crampton 1972). This was further discussed after analysis of a database of 614 carcinogenicity assay results regarding 273 compounds (derived from Soderman, 1982, cited by Schach von Wittenau and Estes, 1983). The need for a second species was highly criticized. The justification for the two species was called "ill-defined," and the choice for two in fact a paradox. Compounds with an inherent property to induce cancer should do so in every species, and thus one species should be sufficient. If a second species would be negative, then the validity for humans would be low, as the finding might be considered species specific.

Schach von Wittenau and Estes (1983) showed that the outcome of mouse studies was similar to that in rats in most cases, and no additional risk assessment could be derived. Most of the compounds listed by them were industrial chemicals with around 10% human pharmaceuticals (including estrogens).

The choice of the second species was therefore identified in the ICH process as an important issue and this was expressed by both industry and regulatory representatives during the first ICH meeting by Drs. Schou and Emerson. Dr. Emerson (from Lilly Research, representing PhRMA) indicated: "*As an animal model*, *the mouse is much less suitable than the rat for reasons frequently enumerated*: *the high background incidence of spontaneous tumours*; *the genetic variability between strains*; *and the small body mass and high rate of metabolism*." The size of the mouse precludes also to measure pharmacodynamic effects during the study (Emerson 1992).

In the same session, Dr. Jens Schou (Danish DKMA, representing EU regulators) indicated: "*I could personally do with only experiments on the rat*, *as mice often create more problems than they add to the prediction*, *especially the problem of liver tumors*" (Schou 1992).

During the early years of ICH, it was decided to build a database of carcinogenicity studies for pharmaceuticals from 1980 on, as from that period most of the carcinogenicity studies were conducted under GLP conditions. A common format was proposed and used in these studies. However, the analysis and evaluation was independent in each region.

Van Oosterhout et al. (1997) described a database built up by the Dutch and German authorities on behalf of the European Economic Community. Not only were the facts important (i.e., the presence, identity, and number of tumors) but also the weight placed on the observations during the evaluative process.

Contrera et al. (1997) published such a database from FDA experience, which included most of the compounds included in the Dutch/German database, but in addition it contained a high number of anonymized compounds under development at that time or terminated at the very end of development.

In the evaluation of these databases, there were two important discussions points:

- The added value of mouse data in case of positive rat data (positive in this case equates with tumors were observed) (see Sect. 5.4.2)
- The value of a positive mouse study when the rat study was negative (see Sect. 5.4.3)

5.4.2 Concordance Between Rat and Mouse Tumor Data

Table 5.1 compares the outcome of several databases with respect to concordance of rat and mouse findings. Schach von Wittenau and Estes (1983) described a concordance of 77% between rats and mice, which is the sum of 120 compounds that are either clearly carcinogenic (86 compounds) or inconclusive (34 compounds with benign tumors only) in rats and mice and 90 compounds that are noncarcinogenic in either species (see also Table 5.1). The conclusion of the authors is that because of the high rate of concordance between rat and mouse, the latter has no added value in risk assessment decisions. Gold et al. (1989) have also published an analysis on a dataset of 392 compounds. The data in Table 5.1 clearly confirm the concordance between rat and mouse (76%).

In the analysis conducted by Van Oosterhout et al. (1997), concordance in rat and mouse outcomes was also in the same range.

Tennant (1993) emphasized the importance of trans-species carcinogenicity, that is, compounds inducing tumors in two species should be classified as having a higher risk in humans than compounds inducing tumors in only one species.

However, from the EU side (Van Oosterhout et al. 1997), it was indicated that this trans-species carcinogenicity could be ascribed primarily to the pharmacological action, while for partial transspecies carcinogenicity the liver was the main common organ, the effect being explained by a direct action on the liver metabolism. This analysis was confirmed recently by Friedrich and Olejniczak (2011) for products reaching the market since 1995–2009.

In line with Tennant (1993), researchers from FDA (Contrera et al. 1997) indicated that carcinogenicity studies in two species are necessary primarily to identify trans-species carcinogens. From this point of view, a reduction to a request for only one species would potentially compromise human safety (Abraham and Reed 2003) (see discussion in Sect. 4.4).

Table 5.1 Rat and mouse carcinogenicity assay concordance

	Schach von Wittenau and Estes (1983)	Van Oosterhout et al. (1997)	Contrera et al. (1997)	NTP (Huff and Haseman 1991)	CPD (Gold et al. 1989)	Friedrich and Olejniczak (2011)
Total compounds with studies in two species	273	181	282	313	392	116
Rat and mouse concordance	210 (77%)	127 (70%)	209 (74.1%)	74.4%	296 (76%)	76
Rat+, mouse+	120 (44%)	34 (19%)	52 (18.4%)		130	32
Rat–, mouse–	90 (33%)	93 (51%)	157 (55.7%)		166	44
Rat and mouse nonconcordance	23%	30%	73 (26%)		96 (24%)	
Rat+, mouse–	26 (9.5%)	41 (23%)	45 (16%)		40	22
Rat–, mouse+	37 (13.5%)	13 (7%)	28 (10%)		56 (14%)	18

The data from Schach von Wittenau and Estes (1983) have been recalculated in the sense that all "inconclusives" are taken as "positives." Some compounds were scored inconclusive as because of inadequate data, but most of them were inconclusive because of the observation of benign tumors only. In the other databases, this differentiation has not been made so strict

5.4.3 The Impact of Mouse-Only Carcinogens

One assessment of the relevancy of the mouse can be derived from the regulatory measures taken on the basis of the outcome of the mouse study, especially when the mouse is the only positive species. In the EU-based paper from Van Oosterhout et al. (1997), this has been studied explicitly in the assessment reports of the two regulatory bodies in Germany, the Bundes Institut für Arzneimittel und Medizin Produkte (BfArM), and the Netherlands, College ter Beoordeling van Geneesmiddelen (CBG, Medicines Evaluation Board). The authors concluded that mouse-only carcinogenicity did not lead to regulatory measures, but it has to be admitted, as was repeated discussed by the EWG, that this conclusion was based on an evaluation of products that were approved for marketing only.

The liver was clearly the most abundant target organ for carcinogenicity (Van Oosterhout et al. 1997; Contrera et al. 1997), confirming data in NTP and CPD databases (Huff and Haseman 1991). In parallel research conducted by FDA, Contrera et al. (1997) discuss two cases, that is, methylphenidate and oxazepam, both inducing liver tumors. Oxazepam induced hepatocellular adenoma and carcinoma after long-term administration in nearly 100% of the animals at the high dose. Hepatoblastoma was observed with a lower frequency. The relevance of mouse liver tumors induced by oxazepam is debated highly (Rauws et al. 1997).

Oxazepam was in this respect similar to phenobarbital. Hepatoblastomas are malignant tumors occurring in children under 3–4 years of age with a different morphology as the hepatocellular adenocarcinoma at a greater age (Frith et al. 1994). It was argued in the EWG by the EU regulators and industry; however, this commonality between mice and human hepatoblastoma is only histopathological and appears not related to the etiology of the carcinogenic response. Hepatoblastoma in humans may occur as a single and early tumor response, while in mice the hepatoblastomas are generally observed with hepatocellular adenomas (Diwan et al. 1994).

Hepatoblastomas were observed also with methylphenidate in mice, as referred to by Contrera et al. (1997). Recent clinical evidence indicates that there is no increased incidence of hepatoblastoma in children, the target population for this medicinal product (Walitza et al. 2010).

In recent years, since the guidance was generated, robust evidence for a mode of action could be sufficient to confirm the safety of compounds inducing mouse liver tumors (Holsapple et al. 2006). The high susceptibility of some mouse strains is reported to be due to a genetic locus (logically called Hcs [hepatocarcinogen sensitivity]) (Drinkwater and Ginsler 1986). Sensitive strains appeared to have a high incidence of spontaneously mutated H-ras oncogenes and are defective in their control of DNA methylation (Counts and Goodman 1994). H-ras oncogenes are considered of limited importance in human cancer (Ozturk 1991).

The relevance of mouse-only tumors was therefore an important discussion point in which different positions became clear between the EU and FDA regulatory authorities: it is clear from the Van Oosterhout's paper that mouse liver tumors

in the EU never led to a decision that these tumors would be relevant to humans, but in FDA experience several undisclosed examples were present where consideration of mouse findings were used in regulatory actions for compounds that did not get marketed.

5.4.4 Compromising Human Safety?

The general public considers animal tests as highly reliable, as this is the basis upon which actions are taken by regulatory authorities with respect to the safety of compounds. However, many tumor responses in rodents have been identified as irrelevant to human by considering the mode of action (Silva Lima and Van der Laan 2000).

Abraham and Reed (2003) have discussed the ICH process on carcinogenicity from a viewpoint of social science and have criticized many of the ICH guidance recommendations. The authors indicate that although it is often claimed that harmonization should accelerate development of important human pharmaceuticals without compromising human safety, they viewed this as not accurate with respect to the ICH carcinogenicity testing guidances. Based on documentary research and interviews, they concluded that the acceleration of drug development is achieved in ICH guidance at the expense of safety standards. As an example, they interpret Dr. Schou's (Schou 1992) published talk as indicating he preferred the approaches prior to the ICH guidance for assessing carcinogenic risk. "Similarly Schou has acknowledged that it is generally agreed that the lifetime carcinogenicity study is the test which gives the optimal answer to the question if a new drug presents a carcinogenic risk." This citation seems to suggest that Schou would be in favor of maintaining rodent life span studies in rats as well as in mice. However, as indicated above, Schou also indicated that he "could live with one species, that is, the rat for this purpose." Clearly, this was in accordance with the discussions that resulted in generation of the new guidance, reducing dependence on 2 lifetime bioassays.

It should be clear from the descriptions above that the eventual S1A, B, and C documents have been discussed thoroughly also from the viewpoint of maintaining human safety.

5.4.5 The Present Text in ICH S1B

Given the numerous, although not unanimous, opinions in the published literature, and the strongly held views of some of the EWG members against the value of the mouse bioassay, why was the mouse testing retained, although modified to allow use of an alternative transgenic mouse rather than the 2-year bioassay?

Insight into this decision can be gained by considering the different interpretations of the databases by the EWG members. As can be derived from the different database overview from EU countries (Van Oosterhout et al. 1997) and FDA

(Contrera et al. 1997), the view of the regulatory authorities on the value of the mouse differed. The conclusion of the EU overview was that no single regulatory action could be attributed to a tumor finding in a mouse carcinogenicity study and that a negative mouse study hardly had contributed to declare a finding in a rat study irrelevant. Thus, in the EU view (and also in the industry view), elimination of the recommendation of testing in mouse was a preferred outcome.

On the contrary, the FDA overview clearly discussed the trans-species findings in line with the classification by Tennant (1993), and two specific product findings were discussed. In the discussion, the FDA experts also referred to several other unpublished cases with mouse-only carcinogens for which no clear mechanism was present at that time, leading to regulatory measures.

A case in point for FDA was the use of a mouse p53 assay that drove the removal of phenolphthalein from the market in the USA. Phenolphthalein received also a negative recommendation from the CPMP in Europe, but this was merely based on its weak genotoxic action in vivo and explicitly not on the data from the mouse p53 assay (CPMP 1997).

In order to avoid a stalemate in the EWG, a compromise was introduced in the guideline to include the mouse, but to give the highest priority to the transgenic mouse models, although at that time the models had not been extensively evaluated for pharmaceutical products. The transgenic mouse models are mentioned as the first option in the text, followed by the full lifetime mouse studies as the second option.

This preference for the transgenic mouse models is not understood easily. In practice, the use of transgenic mouse models appears to have been relatively low, as can be derived from the various reports in the public domain. It may be that the pharmaceutical industry has been reluctant to use these models, as was stated at the time the guidance was created, because of uncertainty in their performance and the interpretation of their results by regulatory authorities. In the development of the guidance, specific models were noted as potentially acceptable, the p53 and Tg.AC mouse proposed by US regulators and the TgRasH2 mouse by the Japanese regulatory EWG members. There was extensive discussion and debate within the EWG as to how these would be used and the value they would add to the cancer risk assessment, but all agreed that this offered the only mutually acceptable path forward at that time.

5.4.6 Further Evaluation of Transgenic Mice

In response to the industry concern about uncertain performance of these assays for pharmaceuticals, ILSI-HESI coordinated an extensive evaluation program of different models, that is (1) the p53-knockout, heterozygous p53 model, (2) the real transgene TgRasH2 with a knock-in of copies of the human RasH2 genes, (3) the transgene Tg.AC based also on a knock-in with multiple copies of a zeta-globulin promoter/v-Ha-Ras gene, and (4) the XPA-p53, a knockout model of a DNA repair

Table 5.2 Performance of individual models for likely human carcinogens and noncarcinogens[a]

Strategy	Positive for carcinogens	Negative for noncarcinogens	Positive for noncarcinogens	Negative for carcinogens	Overall accuracy
$P53^{+/-}$	21	27	1	10	81% (48/59)
$P53^{+/-}$ (G)[b]	16	6	0	4	85% (22/26)
rasH2	21	18	5	7	76% (39/51)
Tg.AC	17	29	10	6	74% (46/62)
$XPA^{-/-}$ and/or[c] $XPA^{-/-}/p53^{+/-}$	7	8	1	2	83% (15/18)

[a] Compounds with IARC classifications 1, 2A, or 2B taken as likely human carcinogens and all IARC classification 3 compounds taken as true noncarcinogens Adapted from Pritchard et al. (2003) and de Vries et al. (2004)

[b] Genotoxic compounds only

[c] Responses detected for one of the two models or both [Reprinted with permission from Storer et al. (2010)]

gene, developed to reflect xeroderma pigmentosa. This was undertaken after the final guidance was published and the results have been published (Storer et al. 2001; Eastin et al. 2001; Usui et al. 2001; van Kreijl et al. 2001) followed by future plans for the evaluation of these models (MacDonald et al. 2004).

At that stage, the FDA reported having assessed about 90 protocols of transgenic mice and two dozens of genetically modified studies (or other alternative assays) had been evaluated. Most of the pharmaceuticals were tested in the $p53^{+/-}$ assay. In the opinion of the FDA, the $p53^{+/-}$ animals are generally appropriate for clearly or equivocal genotoxic drugs. The TgRasH2 model might also been useful to evaluate genotoxic and nongenotoxic drugs.

The Tg.AC model was recommended for dermally applied pharmaceuticals, although also was being evaluated and used for systemically administered compounds.

The EMA has published general conclusions and recommendations (EMA 2004), which were followed by discussion of the state of the art of the individual models. The TgRasH2 as well as the p53 model can be accepted for regulatory purposes, although some individual studies showed equivocal responses. The Tg.AC mouse reacted inconsistently and incompletely to human carcinogens and was therefore restricted for screening the carcinogenicity properties of dermally administered drugs, but could not be recommended for oral studies.

The $XPA^{-/-}$ and the XPA/p53 were declared to be promising, but more data was considered to be needed. One concern was the observed excessive sensitivity of the animals to the effects of p-cresidine and benzo(A)pyrene.

Storer et al. (2010) have reviewed more recently the use of transgenic mice for testing carcinogenicity (Table 5.2). There are a number of carcinogens that are negative in the mouse model, for example, in p53 hemizygous mice. However, this might be rather due to the inclusion of rodent carcinogens in the IARC classification of class 2B (also oxazepam is an IARC 2B possible carcinogen) than reflecting the real human risk of the compounds.

The use of these transgenic models in regulatory testing of pharmaceuticals has been increasing but has not replaced the use of the life span study with mice in the majority of cases. The database published by Friedrich and Olejniczak (2011) covering compounds receiving a marketing authorization between 1995 and 2009 mentions only 1 compound reviewed under the CHMP with a transgenic mouse study.

TgRasH2 mice are recommended as a clear and easy strain to use in assessing carcinogenicity. The TgRasH2 mice were sensitive to PPAR-α-agonists, such as di-ethylhexylphthalate, clofibrate, and WY14643, although clofibrate is believed to be not a human carcinogen (Silva Lima an Van der Laan 2000).

Storer et al. (2010) indicate that the industry is reluctant to use these new models until there is a large historical control dataset like that routinely used in explaining unexpected rare findings in the traditional mouse model, the "devil we know". However, unexpected rare tumors in the transgenic models might be interpreted with more cautiousness. It is this type of conservatism that may be stronger than the willingness to use new approaches in carcinogenicity testing, no matter how resource sparing.

From Table 5.2, it is also clear that the p53 heterozygous mouse is used predominantly to test (equivocally) genotoxic compounds. One of the aspects of this model is that the induction of tumors in the $p53^{+/-}$ mice is associated with a specific loss of the heterozygosity in the tumors, as illustrated in the data with phenolphthalein (Dunnick et al. 1997; Hulla et al. 2001). This effect is described also for other cases as a confirmation that the model is used appropriately.

An overall evaluation of the utility of these assays might be of value after having these recommendations in place and applied for 8–10 years. In recent years, some major PhRMA companies have begun to adopt the use of transgenic models as part of their routine testing paradigm. As a result, the necessary data may soon be available

5.5 Potential Future Directions in Carcinogenicity Testing

5.5.1 Expectations for Future Developments

The current ICH testing guideline S1A as discussed above essentially treats equally all pharmaceuticals that are expected to be administered regularly for 6 months or longer or in a frequent and intermittent manner over a substantial portion of a patient's lifetime. There is presently no acknowledgement for differentiation of carcinogenic risk using a weight-of-evidence approach based on results of short-term studies. On the contrary, current S1A testing guidelines specify additional risk factors, such as structural similarity and previous experience in the chemical class, which would trigger concerns for carcinogenicity testing, even for pharmaceuticals that are used infrequently. The approach for considering factors of additional risk is

reasonable but also could be enhanced by an approach for considering factors that would appropriately reduce concerns.

Any new carcinogenicity testing paradigm would be expected to identify the risk of a pharmaceutical for causing cancer in humans being significant enough to either prevent marketing or to allow marketing but with a meaningful drug warning that would inform a decision regarding the risk–benefit for treatment with that medicine at the prescribed dose. While improving on current capabilities to deliver on these expectations, the hope, furthermore, would be that new approaches would do this faster and/or require fewer animal and human resources.

One approach worth consideration as a near-term future direction for carcinogenicity testing is to introduce a weight-of-evidence approach for assessment of carcinogenic risk (similar as with immunotoxicity testing) and reserve the 2-year testing in rats only for compounds based on real unknown concerns for carcinogenicity without adding substantially to the existing testing requirements.

5.5.2 Prediction of Carcinogenicity Study Outcomes from Noncarcinogenicity Datasets

One proposal considered recently for significant modification to current carcinogenicity testing guidelines is based on the belief that certain risk factors can be used to stratify concern for carcinogenicity. It posits that in the absence of any intended pharmacologic endocrine action, off-target findings in shorter term genotoxicity tests, off-target endocrine perturbation, and off-target histopathologic findings in chronic rat toxicology studies indicative of risk factors for neoplasia, pharmaceuticals of minimal concern could be identified and these criteria used to determine that such compounds need not be tested in a 2-year rat carcinogenicity study (Reddy et al. 2010); Sistare et al. 2011). This proposal is based originally on work by Reddy and subsequently on a proprietary PhRMA database survey of 182 marketed or nonmarketed pharmaceutical development candidates, as well as publicly available data from 78 IARC chemical carcinogens and 8 additional pharmaceuticals withdrawn from the market. The data support the conclusion that pharmaceuticals where a 2-year rat carcinogenicity testing would be expected to add little value to carcinogenicity risk assessment can be identified earlier and a 2-year rat carcinogenicity testing could be supplanted as a test requirement, allowing the results from a carcinogenicity test of a single species—a 6-month transgenic mouse study (see paragraph 5.4.6)—to serve as the only rodent test of carcinogenicity, together with a refined evaluation of chronic and shorter term toxicology tests that identify cancer risk factors. Such exemptions for the conduct of a 2-year rat study should be warranted for certain pharmaceuticals with a strong safety profile in animal and in vitro tests, including a negative outcome in a transgenic mouse carcinogenicity study. Tumorigenic risk potential can be gathered from such a weight-of-evidence approach incorporating both on-target-related pharmacologic effects as well as “off-target” and unanticipated chemical specific actions (generally with an

unknown mechanism). The weight-of-evidence collection of risk factors defined to capture sufficient sensitivity to warrant utility of this proposed negative prediction approach has been outlined with regulatory considerations in work by a consortia of pharmaceutical companies (Sistare et al. 2011). Considering the 182 pharmaceuticals in the PhRMA database, the 78 IARC Group 1 and 2A human chemical carcinogens, plus the 8 additional pharmaceuticals withdrawn from the market due to carcinogenicity concerns, in total 268 chemicals, the proposed criteria correctly identified the need to not conduct a rat carcinogenicity study and conversely those that should be run for further understanding of potential risk. Those ultimately run by failing the exclusion criteria were determined positive for 134 of 148 chemicals yielding 91% test sensitivity. Furthermore, the 14 "misses" (compounds excluded under the criteria but positive in the 2-year rat) were deemed to be of questionable human relevance. For the compounds across the list of 268 chemicals that were deemed to present with human relevant tumorigenicity findings in the rat warranting either withdrawal from marketing, termination of development, or an IARC human carcinogen classification, the criteria demonstrated 100% sensitivity in identifying the need to conduct a 2-year rat carcinogenicity study later shown to be positive. As noted in prior sections above, this latter group consists of only a small number of compounds.

The value of the approach to eliminate the conduct of 2-year rat carcinogenicity studies on pharmaceuticals with no risk factors for carcinogenicity would be the reduction of the time needed to bring important pharmaceuticals to the market to patients in need, the elimination of approximately 600 rats and 400 mice per test compound (if the transgenic mouse model were substituted), and the elimination of approximately $3.75 million in costs associated with the completion and evaluation of each 2-year rat carcinogenicity study. The work by Reddy (2010) and the database survey indicate that approximately 40% of pharmaceuticals would meet the criteria for a 2-year rat carcinogenicity study exemption. Similar results have been reported by a consortia of Japanese pharmaceutical companies using a distinct compound dataset (Hisada et al. 2012).

From these analyses, two critical messages emerge to be embraced in any proposal to guide modifications to future carcinogenicity testing (1) both expected on-target-related excessive pharmacology as well as pharmacology and toxicology unrelated to the primary therapeutic mechanism can yield tumors, so both must be incorporated in the adoption of any new proposed shorter term predictive approaches to modify current testing, and (2) multistep and multi-organ indirect systems biology mechanisms involving sustained disruption and communication across endogenous molecular pathways between tissues will drive nongenotoxic tumorigenesis in rats, and while human relevance is rightfully questioned, the need may exist to diligently investigate such concerns.

In the PhRMA database survey, it was stressed that known endocrine target-related pharmacology is an automatic positive risk factor for the need to investigate rat carcinogenicity and furthermore that any known or discovered disruption of endocrine receptors, of hormone levels, or of local tissue endocrine activity would be considered just cause for the conduct of a 2-year rat carcinogenicity study as a first

step toward identifying the need to understand human relevance. Three categories are discussed below:

- PPAR-gamma agonists
- TSH-inducing mechanisms
- Gastrin elevation

5.5.2.1 Peroxisome Proliferator-Activated Receptor Gamma Agonists

The peroxisome proliferator-activated receptor gamma agonists such as rosiglitazone and pioglitazone and dual alpha/gamma agonists such as muraglitazar and ragaglitazar, for example, would fall into this category based on their known pharmacology to enhance insulin sensitivity. Prior to any experience with this class of compounds, knowledge of mechanism alone would rightfully raise theoretical safety concerns for tumorigenesis that would warrant systematic and thorough experimentation in two species of rodents. Testing in rats revealed human health concerns over bladder tumorigenesis associated with the class in rats, with a benefit-to-risk decision that enabled marketing at the time of product introduction. But even today, questions of human relevance persist (Keiichiro et al. 2011; El-Hage 2005; EMA 2011) and are presently the subject of clinical research during the postmarketing phase (FDA 2011; Lewis et al. 2011).

5.5.2.2 TSH-Inducing Mechanisms

One could argue that, for well-established endocrine mechanisms such as results from liver enzyme induction and subsequent disruption of thyroid signaling, for example, the rat is an inappropriate and overly sensitive model for indirect human thyroid carcinogenesis mechanisms, and so evidence of only such thyroid endocrine-mediated tumors in short-term studies may need not be further investigated with the conduct of a 2-year rat bioassay. Rat liver enzyme inducers have been shown to accelerate turnover of circulating thyroid hormones and elevate TSH levels to chronically stimulate mitogenesis of rat thyroid follicular cells resulting in tumors over the course of a rat's lifetime, but the mechanism is now well accepted to be irrelevant to humans (McClain 1989; Capen 1997, 1999). In fact two recently published independent surveys of carcinogenicity labeling of marketed pharmaceuticals in the United States and in Europe (Alden et al. 2011; Friedrich and Olejniczak 2011) have drawn similar conclusions that **most** treatment-related neoplastic findings seen in rodent carcinogenicity studies are **not** considered relevant for human risk and that significant revision of the carcinogenicity testing paradigm is warranted. When such human irrelevant scenarios are suspected, additional mechanistic assessments such as those described (Silva Lima and Van der Laan 2000; Cohen 2004, 2010) would be expected to improve human carcinogenicity risk assessment and negate the need to conduct a 2-year rat carcinogenicity study. This mode of action framework approach could be deployed early when indirect mechanisms may be suspected from recognized tissue

patterns of histologic changes in chronic studies together with knowledge of pharmacology and specific measurements of tissue molecular and biochemical changes and alterations in hormone levels.

5.5.2.3 Proton Pump Inhibitors

The proton pump inhibitors provide a third categorical example of a pharmacological endocrine mechanism-mediated tumor risk identified in the course of a 2-year rat carcinogenicity tests. The feedback endocrine loop resulting in prolonged hypergastrinemia and stomach tumors in rats following sustained gastric proton pump inhibition and altered local pH has been shown to drive tumorigenesis in rats (Burek et al. 1998). Indeed, while pathologic and primary hypergastrinemia is a viable mechanism for tumorigenesis in humans (Dockray et al. 2005), the levels of gastrin and duration of hypergastrinemia needed to drive tumorigenesis in humans are not reached and sustained to drive such tumors in humans taking proton pump inhibitors. Clinical research conducted in humans treated with proton pump inhibitors settled the contentious issue (Dockray et al. 2005) and allowed an important class of agents to be marketed for the relief of human suffering. This example may demonstrate that 2-year rat carcinogenicity tests can serve a valuable role in identifying risks and can trigger appropriate assessments of interspecies mechanisms. This may involve creative and definitive clinical and nonclinical research approaches to resolve questions of relevance, even including directed human mechanism-based bridging biomarker measurements and imaging approaches.

However, a critical point to acknowledge here is that the redirection of resources to such targeted translational mode of action biomarker applications and clinical research approaches to resolve a hypothetical risk that was reinforced by carcinogenicity testing in rats is more prudent and serves a far greater impact, than routine equivalent investment in 2-year rat carcinogenicity tests on **all** pharmaceuticals. It is reasonable that a pharmaceutical candidate with no identified tumor risk factor signals in chronic rat studies, in vitro genotoxicity studies, or hormonal perturbation studies, and no reasonable hypothetical target-related tumor risks, does not warrant a 2-year testing in the rat.

This proposed approach for small molecules is somewhat analogous to that embraced within ICH guidance S6 for large biologic molecules (ICH 1997). For biologics, the burden is on the sponsor to develop a prudent and thoroughly diligent justification addressing the need or lack thereof for assessing carcinogenic potential following modification of the activity of the drug target by the proposed therapeutic. In some cases, such as for biologic immunomodulators, for example, it is recognized that the pharmacology of such agents is well accepted to result in an enhanced tumor risk in humans so no additional study may be needed and the drug product will be labeled as such, especially since rats are very poor models for immunosuppression-mediated carcinogenicity (Cohen et al. 1991; Bugelski et al. 2010).

A possible approach for small-molecule carcinogenicity testing could be expanded to consider other pharmacologic targets with a likely hypothetical risk for

resulting in carcinogenesis, such as drugs that might target tumor suppressor transcription factors, antiapoptotic proteins, or cell cycle regulators, for example, and not just endocrine target modulation. One view is that a predictable carcinogenic outcome (followed by appropriate labeling) does not warrant the conduct of a full life span carcinogenicity study.

5.5.3 Assessments of the Potential for Emerging Novel Gene Expression Endpoints to Support Carcinogenicity Testing Revisions

On the horizon, advances in molecular biology, genomics, and analytical technologies to expand capabilities and minimize costs of tissue and accessible biomarker measurements have raised hopes that lifetime rodent bioassays could be eliminated and replaced by shorter studies that would more effectively predict human cancer risk and not just rodent cancer risk (reviewed recently in, Guan et al. 2008). Hanahan and Weinberg (2011), in a recent review of the challenges to successful tumor treatment, have framed well the complexity of the challenge that exists, however, for any biomarker based approach to predicting tumor risk from early changes in drug-induced cellular and molecular biology. The authors propose that eight hallmarks of cancer constitute a general organizing principle for understanding the biological capabilities acquired during tumorigenesis—sustained proliferative signaling, evasion of growth suppression, resisting cell death, replicative immortality, angiogenesis, activated invasion and metastasis, reprogramming of energy metabolism, and evasion of immune destruction—while two additional hallmarks expedite or foster the acquisition or function of these eight hallmarks—genome instability and inflammation. Hypothetically, if changes in certain of these hallmarks are conserved across species and across tissues, and a combination of accessible biomarkers, tissue gene expression biomarkers, and histopathologic changes can be measured in samples from organs and tissues of short-term rat studies conducted with known tumorigens to identify the emergence of these hallmarks, then tumor risk may be predictable with a reasonable sensitivity. When fully qualified, the absence of all of these hallmarks might serve as strong indication of the absence of potential carcinogenicity and completely obviate the need for additional testing. It is likely that a reasonable specificity will remain a far more daunting challenge taking such an approach, however. One could surmise that many compounds will provoke several but not all ten hallmarks and elicit microscopically observable confirmatory proliferative changes in shorter term rat studies but not ultimately result in tumors after 2 years of dosing. In the PhRMA database, for example, 38 molecules presented with histologic evidence of risk for potential carcinogenesis in at least one tissue in a rat chronic study, but no tumors were seen after 2 years of dosing (Sistare et al. 2011). Presumably, if tissue biomarkers of several of the ten hallmarks could be measured confidently, they would be present

in these tissues presenting with that histology, but this hypothesis remains to be evaluated. In addition, one could imagine a case involving the 14 false negatives identified by the PhRMA group, assuming these are legitimate consistent and reproducible false negatives, where the novel tissue biomarkers might be positive and therefore outperform the lack of histologic findings at the 6-month time point. For any new such testing paradigm incorporating the measurement of potential novel tissue biomarker endpoints to be accepted for regulatory decision making, a testing strategy would be needed using a comprehensive approach with a balanced set of true positives and true negatives, building upon the historical test data and critical compounds identified in the historical database to represent legitimate regulatory concerns.

Attempts have been made by several groups to identify and establish reproducible broadly predictive tissue gene expression biomarker signatures qualified for predicting drug-induced carcinogenicity potential (Kramer et al. 2004; Nie et al. 2006; Ellinger-Ziegelbauer et al. 2008; Fielden et al. 2007; Uehara et al. 2008). The expectation would be that gene expression changes in the tumor target tissue in a short duration rat study would reflect several of the ten earliest hallmarks of biological change associated with tumorigenesis and thereby precede and predict tumor development seen in a 2-year rat study. In theory the gene expression signature biomarkers should be independent of drug mechanism and broadly applicable across drug classes. The Predictive Safety Testing Consortium evaluated several published gene expression signatures across a number of independently gathered sample sets, focusing on nongenotoxic hepatocarcinogenicity prediction. Initial interlaboratory results were encouraging (Fielden et al. 2008). Subsequent follow-up research efforts by the consortium focused on the performance evaluation of a 22-gene signature using a single PCR-based platform across a diverse set of samples from livers of rats treated with an independent set of 66 rat liver nongenotoxic hepatocarcinogens or nonhepatocarcinogens collected from several laboratories. The authors reported rather low 67% sensitivity and 59% specificity, noting however and in agreement with Auerbach et al. (2010) that matching the strain of rat and the duration of dosing of the study samples to the test set samples used to derive the signature may be critically important study protocol criteria to consider for any further evaluations of gene expression signature prediction performance.

Recently, Uehara et al. (2011) report 99% sensitivity and 97% specificity for rat hepatocarcinogenicity prediction using training set data derived from an established large-scale toxicogenomics database known as TG-GATEs (Genomics-Assisted Toxicity Evaluation System developed by the Toxicogenomics Project in Japan). An independent assessment of a signature by the authors was conducted using publicly available gene expression data, obtaining 100% sensitivity and 89% specificity. However, while the data were generated independently, many of the same compounds appear in both the training set and the independent public database test set and call into question the extent of concordance reported. Moreover, the value in predicting hepatocellular cancer is questionable. Clearly, if gene expression endpoints are to be proposed as tissue biomarkers to be added

to a weight-of-evidence approach to reducing 2-year rat carcinogenicity testing, many questions remain to be answered and systematically evaluated. The authors conclude that the approach may be useful now for internal decision making for screening for potential for hepatocarcinogenicity of compounds early in drug development and it is likely premature to include such assessment in regulatory studies for regulatory decisions. However, even this limited application could be of minimal utility and value by the generally low concern for rodent hepatocarcinogens as discussed above.

5.5.4 Developments in Epigenetics Including Noncoding RNAs

In parallel to the maturation of toxicogenomic strategies for investigating mechanisms and biomarkers of nongenotoxic carcinogenesis, the role of epigenetic mechanisms is beginning to gain attention. Transcriptomic mRNA profiles derived from toxicogenomic approaches reflect the dynamic interplay between a diverse range of transcription factors and epigenetic regulatory proteins. Epigenetics describes heritable changes in gene function that occur in the absence of a change in DNA sequence. Epigenetic modifications of the genome include methylation of DNA at cytosine residues and posttranslational modification of histone proteins that package DNA into chromatin. Noncoding RNAs and higher-order levels of chromatin structure also contribute to the epigenetic regulation of gene expression. Numerous chromatin-modifying proteins contribute to the establishment and maintenance of combinatorial epigenetic signatures that functionally organize the genome. The epigenome is subject to short-term dynamic changes (e.g., during DNA transactions including replication, repair, recombination, and transcription) but can also retain stable long-lasting modifications that form the molecular basis for developmental stage and cell type-specific patterns of gene expression. Recent insights into the molecular and biochemical mechanisms that enable cells to read, write, and erase epigenetic codes have revealed a close association between epigenetic changes and the predisposition to, and development of, a wide range of diseases including cancer (Portela and Esteller 2010).

Emerging data suggest that epigenetic perturbations may also be involved in the adverse effects associated with some drugs and toxicants, including certain classes of nongenotoxic carcinogens (Marlowe et al. 2009; Lempiäinen et al. 2012a, b). Importantly, the stable propagation of epigenetic modifications through mitosis and cell division provides a mechanistic basis for long-lasting xenobiotic-mediated cellular perturbations including carcinogenesis. In contrast to the classical model of multistage carcinogenesis, in which successive genetic changes result in initiation, promotion, and progression, the epigenetic progenitor model of cancer (Feinberg et al. 2006) postulates that a combination of epigenetic and genetic changes contributes to each stage. Furthermore, epigenetic modifications can also contribute directly to genomic instability as exemplified by point mutations associated with the spontaneous deamination of 5-methylcytosine. The recent expansion of the

mammalian DNA methylome to include three additional modified DNA bases (5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxylcytosine) that are regulated via both epigenetic and DNA repair pathways (Wu and Zhang 2011) suggests an increasing importance of considering genetic–epigenetic interactions during cancer risk assessment.

The potential importance of epigenetic mechanisms of nongenotoxic carcinogenesis has been a key driver for the Innovative Medicines Initiative MARCAR Consortium initiative (2010–2014; http://www.imi-marcar.eu) whose goal is to explore the utility of integrating novel molecular profiling technologies (including DNA methylation, histone modifications, mRNA, noncoding RNA, and phosphoproteins) for mechanistic insight and early biomarkers in rodent models for nongenotoxic carcinogen. In parallel to investigating early mechanism-based markers, the utility of these integrated molecular profiling technologies for molecular classification of rodent tumors (spontaneous vs. drug-induced) is also being explored. MARCAR's initial focus has been on epigenetic mechanisms and biomarkers for well-characterized rodent hepatocarcinogens, although this approach is now being extended to non-liver non-genotoxic carcinogens. The mechanistic basis for early non-genotoxic carcinogen-induced changes in specific epigenetic marks and their potential relevance to nongenotoxic carcinogenesis is being explored using (1) transgenic mouse models (knockout; humanized) for key nuclear receptors and cancer signaling pathways, (2) liver tumor-sensitive and resistant mouse strains, (3) rodent and human liver-derived parenchyme–mesenchyme coculture models, and (4) oxidative stress reporter mice. One of the most promising novel MARCAR non-genotoxic carcinogenesis biomarkers to date is a cluster of long noncoding RNAs and microRNAs that have previously been associated with stem cell pluripotency in mice and various neoplasms in humans (Lempiäinen et al. 2012a). Non-genotoxic carcinogen-mediated induction of these ncRNA biomarkers in mouse liver is dependent both on the constitutive androstane receptor and beta-catenin pathways and is also maintained in non-genotoxic carcinogen-promoted mouse liver tumors (Lempiäinen et al. 2012b). The sensitivity, specificity, dose response, and reversibility of candidate early non-genotoxic carcinogenesis biomarkers resulting from these studies is subsequently being assessed in industry-relevant mouse and rat strains using a panel of known genotoxic and nongenotoxic carcinogens versus appropriate noncarcinogen controls. Of particular interest would be to explore whether novel early epigenetic and/or noncoding RNA non-genotoxic carcinogenesis biomarkers could enhance the prediction of positive rodent bioassay outcomes.

Challenges in the biological interpretation of epigenomic mechanisms and biomarkers include species, tissue, and cell type specificity combined with dynamic changes associated with age, diet, and xenobiotic exposure (Goodman et al. 2010; Lempiäinen et al. 2012a). A major knowledge gap is thus to elucidate the dynamic range of normal epigenetic patterns of variation and to define thresholds above which an epigenetic perturbation might be deemed to be adverse. MARCAR has recently made significant progress in the evaluation of epigenome dynamics in preclinical animal models. Tissue-specific DNA methylomes for mouse liver and kidney have been characterized at the genome-wide level in the context of mechanisms

and early biomarkers for nongenotoxic carcinogenesis and reveal tissue-specific xenobiotic-induced perturbations of DNA methylation at a limited number of gene promoters following chronic exposure to the rodent hepatocarcinogen phenobarbital (Lempiainen et al. 2011). MARCAR is currently performing additional studies to further define tissue, age, gender, strain, and species differences in epigenomes, as well as the functional significance of perturbation by xenobiotics. Central to these efforts will be to ensure the robust phenotypic anchoring of both novel transcriptomic and epigenomic predictive biomarkers to adverse histopathological outcomes (Lempiäinen et al. 2012a).

As a note of caution, however, it needs to be emphasized as in earlier parts of this chapter that rodent carcinogenesis highly overpredicts human cancer risk. This was recognized early on, to the extent that the EU regulators proposed elimination of mouse bioassays. The application of new approaches needs to carefully be assessed to predict real risk to humans, rather than raise unsubstantiated, nonvalid concerns for humans. The future of carcinogenicity testing should not replicate/reinforce errors committed in the past.

References

Abraham J, Reed T (2003) Reshaping the carcinogenic risk assessment of medicines: international harmonization for drug safety, industry/regulator efficiency or both? Social Sci Med 57: 195–204

Alden CL, Lynn A, Bourdeau A, Morton D, Sistare FD et al (2011) A critical review of the effectiveness of rodent pharmaceutical carcinogenesis testing in predicting for human risk. Vet Pathol 48(3):772–84

Auerbach SS, Shah RR, Mav D, Smith CS, Walker NJ, Vallant MK, Boorman GA, Irwin RD (2010) Predicting the hepatocarcinogenic potential of alkenylbenzene flavoring agents using toxicogenomics and machine learning. Toxicol Appl Pharmacol 243:300–314

Bergman K, Olofsson I, Sjöberg P (1998) Dose selection for carcinogenicity studies of pharmaceuticals: systemic exposure to phenacetin at carcinogenic dosage in the rat. Regul Toxicol Pharmacol 28:226–229

Bolognani F, Lempiainen H, Arne M, Philippe C, Dubost V, Luisier R, Thomson J, Unterberger E, Hahne F, Brasa S, Zollinger T, Marcellin M, Marlowe J, Oakeley E, Schubeler D, Schwarz M, Meehan R, Ellinger-Ziegelbauer H, Wolf R, Theil D, Heard D, Moulin P, Grenet O, Moggs J, Terranova R (2012) Comprehensive mapping of transcriptional and epigenetic perturbations in vivo identifies noncoding RNAs as novel biomarkers for liver tumor promotion, 82. The Toxicologist CD, Society of Toxicology, San Francisco, CA, 11 March 2012, p 126

Bugelski PJ, Volk A, Walker MR, Krayer JH, Martin P, Descotes J (2010) Critical review of preclinical approaches to evaluate the potential of immunosuppressive drugs to influence human neoplasia. Int J Toxicol 29:435–466

Burek JD, Patrick DH, Gerson RJ (1998) Weight of biological evidence for assessing carcinogenicity. In: Grice HC, Cimina JL (eds) Carcinogenicity. Springer, New York, pp 83–95

Capen CC (1997) Mechanistic data and risk assessment of selected toxic end points of the thyroid gland. Toxicol Pathol 25:39–48

Capen CC (1999) Thyroid and parathyroid toxicology, mechanisms of toxicity: thyroid follicular cells. In: Harvey PW, Rush KC, Cockburn A (eds) Endocrine and hormonal toxicology. Wiley, New York, pp 42–47

Cohen SM, Purtilo DT, Ellwein LB (1991) Ideas in pathology. Pivotal role of increased cell proliferation in human carcinogenesis. Mod Pathol 4:371–82

Cohen SM (2004) Human carcinogenic risk evaluation: an alternative approach to the two-year rodent bioassay. Toxicol Sci 80:225–229

Cohen SM (2010) Evaluation of possible carcinogenic risk to humans based on liver tumors in rodent assays: the two-year bioassay is no longer necessary. Toxicol Pathol 38:487–501

Contrera JF, Jacobs AC, Prasanna HR, Mehta M, Schmidt WJ, DeGeorge JJ (1995) A systemic exposure-based alternative to the maximum tolerated dose for carcinogenicity studies of human therapeutics. J Am Coll Toxicol 14:1–10

Contrera JF, Jacobs AC, DeGeorge JJ (1997) Carcinogenicity testing and the evaluation of regulatory requirements for pharmaceuticals. Regul Toxicol Pharmacol 25:130–145

Counts JL, Goodman JI (1994) Hypomethylation of DNA: an epigenetic mechanism involved in tumor promotion. Mol Carcinog 11(4):185–8

De Vries A, van Steeg H, Opperhuizen A (2004) Transgenic mice as alternatives in carcinogenicity testing: Current Status. RIVM report 340700001. Available at http://rivm.openrepository.com/rivm/bitstream/10029/8911/1/340700001.pdf. Accessed 27 Nov 2012

Diwan BA, Ward JM, Kurata Y, Rice JM (1994) Dissimilar frequency of hepatoblastomas and hepatic cystadenomas and adenocarcinomas arising in hepatocellular neoplasms of D2B6F1 mice initiated with N-nitrosodiethylamine and subsequently given Arachlor-1254, dichlorodiphenyltrichloroethane or phenobarbital. Toxicol Pathol 22:430–439

Dockray G, Dimaline R, Varro A (2005) Gastrin: old hormone, new functions. Pflugers Arch Eur J Physiol 449:344–355

Drinkwater NR, Ginsler JJ (1986) Genetic control of hepatocarcinogenesis in C57BL/6J and C3H/HeJ inbred mice. Carcinogenesis 10:1701–1707

Dunnick JK, Hardisty JF, Herbert RA, Seely JC, Furedi-Machacek EM, Foley JF, Lacks GD, Stasiewicz S, French JE (1997) Phenolphthalein induces thymic lymphomas accompanied by loss of the p53 wild type allele in heterozygous p53-deficient (+/–) mice. Toxicol Pathol 25(6):533–40

Eastin WC, Mennear JH, Tennant RW, Stoll RE, Branstetter DG, Bucher JR, McCullough B, Binder RL, Spalding JW, Mahler JF (2001) Tg.AC genetically altered mouse: assay working group overview of available data. Toxicol Pathol 29:60–80

El-Hage J (2005) Peroxisome proliferator-activated receptor agonists: carcinogenicity findings and regulatory recommendations. In: International arthrosclerosis society symposium on PPAR, Monte Carlo

Ellinger-Ziegelbauer H, Gmuender H, Bandenburg A, Ahr HJ (2008) Prediction of a carcinogenic potential of rat hepatocarcinogens using toxicogenomics analysis of short-term in vivo studies. Mutat Res 637:23–39

Emerson JL (1992) High dose selection in the design of studies to evaluate the carcinogenic potential of pharmaceuticals: industry perspectives. In: D'Arcy PF, Harron DWG (eds) Proceedings of the second international conference on harmonisation, Brussels, 1991. Queen's University, Belfast, pp 202–208

European Medicines Agency (1997) Position paper on the genotoxic and carcinogenic potential of phenolphthalein. CPMP/818/97, http://www.ema.europa.eu/docs/en_GB/document_library/Position_statement/2009/09/WC500003148.pdf. Accessed 6 July 2012

European Medicines Agency (2004) CPMP SWP Conclusions and Recommendations on the use of genetically modified animal models for carcinogenicity assessment. CPMP/SWP/2592/02 Rev.1 … http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003257.pdf. Accessed 6 July 2012

European Medicines Agency (2011) European Medicines Agency recommends new contra-indications and warnings for pioglitazone to reduce small increased risk of bladder cancer: benefit-risk balance remains positive in a limited population of type 2 diabetics. http://www.ema.europa.eu/docs/en_GB/document_library/Press_release/2011/07/WC500109176.pdf

FDA (2008) Safety testing of drug metabolites. Food and Drug Administration, Washington, DC. http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm06501. Accessed 14 Feb 2008

FDA (2011) FDA drug safety communication: update to ongoing safety review of Actos (pioglitazone) and increased risk of bladder cancer. Food and Drug Administration, Washington, DC. http://www.fda.gov/Drugs/DrugSafety/ucm259150.htm. Accessed 15 June 2011

Feinberg AP, Ohlsson R, Henikoff S (2006) The epigenetic progenitor origin of human cancer. Nat Rev Genet 7:21–33

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

Fielden MR, Nie A, McMillian M, Elangbam CS, Trela BA, Yang Y, Dunn RT II, Dragan Y, Fransson-Stehen R, Bogdanffy M et al (2008) Interlaboratory evaluation of genomic signatures for predicting carcinogenicity in the rat. Toxicol Sci 103:28–34

Friedrich A, Olejniczak K (2011) Evaluation of carcinogenicity studies of medicinal products for human use authorized via the European centralized procedure (1995–2009). Regul Toxicol Pharmacol 60:225–248

Frith CH, Ward JM, Turusov VS (1994) Tumours of the liver. IARC Sci Publ. 111:223–269

Gold LS, Bernstein L, Magaw R, Slone TH (1989) Interspecies extrapolation in carcinogenesis: prediction between rats and mice. Environ Health Perspect 81:211–219

Goodman JI, Augustine KA, Cunnningham ML, Dixon D, Dragan YP, Falls JG, Rasoulpour RJ, Sills RC, Storer RD, Wolf DC et al (2010) What do we need to know prior to thinking about incorporating an epigenetic evaluation into safety assessments? Toxicol Sci 116:375–381

Grasso P, Crampton RF (1972) The value of the mouse in carcinogenicity testing. Food Cosmet Toxicol 10(3):418–26

Guan P, Olaharski A, Fielden M, Roome N, Dragan Y, Sina J (2008) Biomarkers of carcinogenicity and their roles in drug discovery and development. Expert Rev Clin Pharmacol 1:759–771

Hanahan D, Weinberg RA (2011) Hallmarks of cancer: the next generation. Cell 144:646–674

Hayashi Y (1992) Regulatory perspectives on issues in carcinogenicity tests: highest dose based in exposure and kinetics. In: D'Arcy PF, Harron DWG (eds) Proceedings of the second international conference on harmonisation, Brussels, 1991. Queen's University, Belfast, pp 200–202

Hernandez LG, Slob W, Van Steeg H, Van Benthem J (2011) Can carcinogenicity potency be predicted from in vivo genotoxicity data? A meta-analysis of historical data. Environ Mol Mutagen. doi:10.1002/em.20651

Hisada S, Wakata A, Sasaki S, Sawada S, Kudo S, Wato E, Kumazawa T, Moriyama K, Mishima M, Kasahara Y, Kito N, Inoue K, Nakamura K (2012) Japanese questionnaire data survey on predictivity for negative tumor outcomes in two-year rat carcinogenicity studies. In: Poster at SOT, San Francisco, CA, 11 March 2012

Holsapple MP, Pitot HC, Cohen SH, Boobis AR, Klaunig JE, Pastoor T, Dellarco VL, Dragan YP (2006) Mode of action in relevance of rodent liver tumors to human cancer risk. Toxicol Sci 89:51–56

Huff J, Haseman J (1991) Long-term chemical carcinogenesis experiments for identifying potential human cancer hazards: collective database of the National Cancer Institute and National Toxicology Program (1976–1991). Environ Health Perspect 96:23–31

Hulla JE, French JE, Dunnick JK (2001) Chromosome 11 loss from thymic lymphomas induced in heterozygous Trp53 mice by phenolphthalein. Toxicol Sci 60(2):264–70

ICH (1992) Session III (ii) workshop safety. In: D'Arcy PF, Harron DWG (eds) Proceedings of the first international conference on harmonisation, Brussels 1991. Queen's University, Belfast, pp 183–251 (590p)

International Conference on Harmonization (1994) Draft guideline on dose selection for carcinogenicity studies of pharmaceuticals; availability. Fed Reg 59:9752–9755

International Conference on Harmonisation (1997) ICH guideline S6: preclinical safety evaluation of biotechnology-derived pharmaceuticals. July 1997, Revised 2011. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Safety/S6_R1/Step4/S6_R1_Guideline.pdf. Accessed 6 July 2012

International Conference on Harmonisation of Technical Requirements for Pharmaceuticals for Human Use ICH S5 (2000) http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Safety/S5_R2/Step4/S5_R2__Guideline.pdf. Accessed 6 July 2012

International Conference on Harmonisation of Technical Requirements for Pharmaceuticals for Human Use ICHM3(R2) (2009) http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Multidisciplinary/M3_R2/Step4/M3_R2__Guideline.pdf. Accessed 22 Feb 2012

Keiichiro S, Yasuyuki A, Hitoshi K, Tanakamarua Z-Y, Nagaia H, Baronb D, Yamamotoa M (2011) Suppressive effects of acid-forming diet against the tumorigenic potential of pioglitazone hydrochloride in the urinary bladder of male rats. Toxicol Appl Pharmacol 251: 234–244

Kramer JA, Curtiss SW, Kolaja KL, Alden CL, Blomme EA, Curtiss WC, Davila JC, Jackson CJ, Bunch RT (2004) Acute molecular markers of rodent hepatic carcinogenesis identified by transcription profiling. Chem Res Toxicol 17:463–470

Lempiainen H, Muller A, Brasa S, Teo SS, Roloff TM, Morawiec L, Zamurovic N, Vicart A, Funhoff E, Couttet P, Schubeler D, Grenet O, Marlowe J, Moggs J, Terranova R (2011) Phenobarbital mediates an epigenetic switch at the constitutive androstane receptor (CAR) target gene Cyp2b10 in the liver of B6C3F1 mice. PLoS One 6:e18216

Lempiäinen H, Luisier R, Müller A, Marc P, Heard DJ, Bolognani F, Moulin P, Couttet P, Grenet O, Marlowe J, Moggs J, Terranova R (2012a) Epigenomics – impact for drug safety sciences. In: Sahu S (ed) Toxicology and epigenetics. Wiley, New York

Lempiäinen H, Couttet P, Bolognani F, Müller A, Dubost V, Luisier R, Del Rio Espinola A, Vitry V, Unterberger E, Thomson JP, Treindl F, Metzger U, Wrzodek C, Hahne F, Zollinger T, Brasa S, Kalteis M, Marcellin M, Giudicelli F, Braeuning A, Morawiec L, Zamurovic N, Längle U, Scheer N, Schübeler D, Goodman J, Chibout SD, Marlowe J, Theil D, Heard DJ,Grenet O, Zell A, Templin MF, Meehan RR, Wolf CR, Elcombe CR, Schwarz M, Moulin P, Terranova R, Moggs JG (2012b) Identification of Dlk1-Dio3 imprinted gene cluster non-coding RNAs as novel candidate biomarkers for liver tumor promotion. Toxicol [Epub ahead of print]

Lewis JD, Ferrara A, Peng T, Hedderson M, Bilker WB, Quesenberry CP, Vaughn DJ, Nessel L, Selby J, Strom BL (2011) Risk of bladder cancer among diabetic patients treated with pioglitazone interim report of a longitudinal cohort study. Diabetes Care 34:916–922

MacDonald J, French JE, Gerson R, Goodman J, Jacobs A, Kasper P, Keller D, Lavin A, Long G, McCullough B, Sistare F, Storer R, Van der Laan JW (2004) The utility of transgenic mouse assays for identifying human carcinogens – a basic understanding and path forward. Toxicol Sci 77:188–194

Marlowe J, Teo SS, Pognan F, Chibout SD, Moggs JG (2009) Mapping the epigenome – impact for toxicology. EXS Experientia 99:259–88

McClain RM (1989) The significance of hepatic microsomal enzyme induction and altered thyroid function in rats: implications for thyroid gland neoplasia. Toxicol Pathol 17:294–306

Monro A (1994) Utility of two rodent species: some arguments for and against. In: D'Arcy PF, Harron DWG (eds) Proceedings of the second international conference on harmonisation, Orlando, 1993. Queen's University, Belfast, pp 293–300

Monro A, Mordenti J (1995) Expression of exposure in negative carcinogenicity studies: dose/body weight, dose/body surface area, or plasma concentrations? Toxicol Pathol 23: 187–198

Nie AY, McMillian M, Parker JB, Leone A, Bryant S, Yieh L, Bittner A, Nelson J, Carmen A, Wan J et al (2006) Predictive toxicogenomics approaches reveal underlying molecular mechanisms of nongenotoxic carcinogenicity. Mol Carcinog 45:914–933

Ozturk M (1991) p53 mutation in hepatocellular carcinoma after aflatoxin exposure. Lancet 338: 1356–9

Portela A, Esteller M (2010) Epigenetic modifications and human disease. Nat Biotechnol 28: 1057–1068

Pritchard JB, French JE, Davis BJ, Haseman JK (2003) The role of transgenic mouse models in carcinogen identification. Environm Health Perspect 111:444–454

Rauws AG, de Waal EJ, van der Laan JW (1997) Sense and non-sense in toxicity assessment of medicinal products. Adv Drug Res 30:15–72

Reddy MV, Sistare FD, Christensen JS, DeLuca JG, Wollenberg GK, DeGeorge JJ (2010) An Evaluation of Chronic 6- and 12-Month Rat Toxicology Studies as Predictors of 2-Year Tumor Outcome. Vet Pathol 47:614–629

Schach von Wittenau M, Estes PC (1983) The redundancy of mouse carcinogenicity bioassays. Fundam Appl Toxicol 3(6):631–9

Schou JS (1992) Predictive value of carcinogenicity studies for drug safety: with special emphasis on survivability and study duration. In: D'Arcy PF, Harron DWG (eds) Proceedings of the second international conference on harmonisation, Brussels 1991. Queen's University, Belfast, pp 208–212

Silva Lima B, van der Laan JW (2000) Mechanisms of nongenotoxic carcinogenesis and assessment of the human hazard. Regul Toxicol Pharmacol 32:135–143

Sistare FD, Morton D, Alden C, Christensen J, Keller D et al (2011) An analysis of pharmaceutical experience with decades of rat carcinogenicity testing: support for a proposal to modify current regulatory guidelines. Toxicol Pathol 39:716–744

Storer RD, French JE, Haseman J, Hajian G, LeGrand EK, Long GG, Mixson LA, Ochoa R, Sagartz JE, Soper KA (2001) P53+/− hemizygous knockout mouse: overview of available data. Toxicol Pathol 29(Suppl):30–50

Storer RD, Sistare FD, Reddy MV, DeGeorge JJ (2010) An industry perspective on the utility of short-term carcinogenicity testing in transgenic mice in pharmaceutical development. Toxicol Pathol 38:51–61

Tennant RW (1993) Stratification of rodent carcinogenicity bioassay results to reflect relative human hazard. Mutat Res 286:111–8

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

Uehara T, Minowa Y, Morikawa Y, Kondo C, Maruyama T, Kato I, Nakatsu N, Igarashi Y, Ono A, Hayashi H, Mitsumori K, Yamada H, Ohno Y, Urushidani T (2011) Prediction model of potential hepatocarcinogenicity of rat hepatocarcinogens using a large-scale toxicogenomics database. Toxicol Appl Pharmacol 255:297–306

Usui T, Mutai M, Hisada S, Takoaka M, Soper KA, McCullough B, Alden C (2001) CB6F1-rasH2 mouse: overview of available data. Toxicol Pathol 29(Suppl):90–108

van Kreijl CF, McAnulty PA, Beems RB, Vynckier A, van Steeg H, Fransson-Steen R, Alden CL, Forster R, van der Laan JW, Vandenberghe J (2001) Xpa and Xpa/p53+/− knockout mice: overview of available data. Toxicol Pathol 29(Suppl):117–27

Van Oosterhout JPJ, Van der Laan JW, De Waal EJ, Olejniczak K, Hilgenfeld M, Schmidt V, Bass R (1997) The Utility of two rodent species in carcinogenic risk assessment of pharmaceuticals in Europe. Regul Toxicol Pharmacol 25:6–17

Walitza S, Kämpf K, Oli RG, Warnke A, Gerlach M, Stopper H (2010) Prospective follow-up studies found no chromosomal mutagenicity of methylphenidate therapy in ADHD affected children. Toxicol Lett 193:4–8

WHO (1961) Technical Report Series 220. Evaluation of the carcinogenic hazards of food additives. 5th Report of the Joint FAO/WHO Expert Committee on Food Additives

WHO (1969) Technical Report Series 426. Principles for the testing and evaluation of drugs for carcinogenicity. Report from a WHO Scientific Group'

WHO, GLOBOCAN (2008) (IARC) Section of Cancer Information http://globocan.iarc.fr/factsheets/populations/factsheet.asp?uno=900 (checked 27/11/2012)

Wu H, Zhang Y (2011) Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation. Genes Dev 25:2436–2452

Chapter 6
The Evolution, Scientific Reasoning and Use of ICH S2 Guidelines for Genotoxicity Testing of Pharmaceuticals

Lutz Müller, David Tweats, Sheila Galloway, and Makoto Hayashi

Abstract Two ICH guidances on genotoxicity (ICH S2A and ICH S2B) have been put into practice in the ICH regions in 1995 and 1997. At the end of 2011, these were replaced by the revised single ICH S2(R1) guidance. In the context of safety testing of pharmaceuticals, genotoxicity testing is mainly associated with the goal to remove potentially genotoxic carcinogens early in the process of drug development, and this goal requires a battery of different tests to address the various genotoxic mechanisms involved in carcinogenesis. In the years of use of the first ICH S2 guidelines, it has been recognised that the extreme focus on sensitivity for in vitro genotoxicity tests, as well as general improvements in various test systems, requires a revision of the principles of S2A and S2B. Thus, an ICH expert working group was established, which merged the two ICH S2 guidances into one, the ICH S2(R1) guidance. Essential changes in the way to conduct genotoxicity testing of pharmaceuticals include a reduction in the top concentrations used for testing of pharmaceutical candidate compounds in in vitro genotoxicity tests and an option to omit in vitro genotoxicity tests with mammalian cells in vitro from the test battery with inclusion of a more comprehensive in vivo testing. The revised ICH S2(R1) will enable a better risk-based assessment for genotoxicity of pharmaceuticals.

L. Müller (✉)
Non-clinical Safety, F. Hoffmann-La Roche Ltd., Basel, Switzerland
e-mail: Lutz.Mueller@roche.com

D. Tweats
The Genetics Department, The School of Medicine, University of Swansea, Swansea, UK

S. Galloway
Merck Research Laboratories , West Point, PA, USA

M. Hayashi
Biosafety Research Center, Foods, Drugs and Pesticides (BSRC),
Shioshinden, Iwata, Shizuoka, Japan

J.W. van der Laan and J.J. DeGeorge (eds.), *Global Approach in Safety Testing*, AAPS Advances in the Pharmaceutical Sciences Series 5, DOI 10.1007/978-1-4614-5950-7_6,

6.1 Genotoxicity as a Discipline of Regulatory Safety Testing

6.1.1 Historical Overview of Regulatory Genotoxicity Testing Guidelines

In the late 1940s, it was Charlotte Auerbach and colleagues who demonstrated that chemicals could be mutagens based on changes in the genetic information carried in the nucleus of each individual cell of an organism (Auerbach and Robson 1946). Both damage to the germ line that could cause heritable disease and genetic alterations to individuals via somatic DNA damage were in the research focus (DHEW 1977; Meselson 1971; Wassom 1989). The concern that exposure to environmental chemicals could introduce deleterious heritable alterations in the DNA of human beings in the environment led to the formation of the Environmental Mutagen Society in 1969 (Wassom 1989) and to the introduction of requirements for testing for mutagenic properties of chemicals in the 1970s. In this context, the US Toxic Substances Control Act of 1976 specifically required the US Environmental Protection Agency (EPA) to establish standards for the assessment of health and environmental effects associated with mutagenesis (US TSCA 1976). During the early period of mutation research and regulatory actions, the primary focus was on the potential of any chemical to induce germ-line mutations and the development of appropriate testing methodologies (Ehling et al. 1978; Meselson 1971).

The thinking of the field at this key stage is illustrated by a key report of the department-wide working group of the then US Department of Health, Education and Welfare (DHEW) issued in 1977 (DHEW 1977). This working group, formed by the DHEW Committee to Coordinate Toxicology and Related Programs, Subcommittee on Environmental Mutagenesis, was established in 1974. The intent was:

> … to aid officials of regulatory agencies who have the responsibility for deciding (1) advisability of promulgating test requirements for mutagenicity at the present time under any of their legislative authorities; (2) the appropriateness of mutagenicity tests for a wide range of product use and exposure categories; and (3) the reliability and interpretation of data from mutagenicity tests developed on substances of commerce within their regulatory purview in spite of the absence of formal testing requirements.

This report, entitled ‘Approaches to Determining the Mutagenic Properties of Chemicals: Risk to Future Generations’, emphasised the need for a quantitative assessment of the risk of heritable damage to the human germ line. Additionally, the association of mutagenesis with other toxicological endpoints such as carcinogenesis, teratogenesis and aging was also noted.

In the mid-1970s, the landmark publications of Bruce Ames’s team on the detection of carcinogens as mutagens, based on an analysis of 300 chemicals, demonstrated a strong correlation of mutagenic activity in *Salmonella* with animal carcinogenicity (Ames et al. 1975; Maron and Ames 1983; McCann et al. 1975). These papers generated great enthusiasm that inexpensive in vitro mutagenesis screening tests could be used to identify chemical carcinogens, and hence, control

of exposure to such agents could potentially lower the human tumour burden. As regulatory guidelines were implemented during the 1970s and 1980s, there was a shift in focus from concern over germ-line mutagenesis to control of chemical carcinogens (MacGregor 1994). Although these early results in *Salmonella* were highly promising, it was already recognised at that time that mutations could arise by multiple mechanisms, some of which would not be detected in a reversion assay of genes concerned with amino acid biosynthesis, such as the *Salmonella typhimurium his*⁻ reversion test. In particular, chromosomal interchanges, DNA strand breaks and large chromosomal deletions, all characteristic of damage induced by ionising radiation, which was one of the environmental mutagens of most concern during this period, are not efficiently detected in the Ames assay. Thus, an in vitro and in vivo test battery was devised that would detect the major classes of damage known to result in heritable mutations (NRC 1983). These concepts underlie the batteries currently still in use (Brusick 1987). The types of lesions expected to be detected by the test systems most commonly used for mutagenesis screening at the present time are in line with our knowledge about the types of lesions involved in modifying the activity of oncogene products and tumour suppressor gene products. In the meantime, changes in such genes are widely accepted to be associated with cancer risk.

Guidelines for testing environmental chemicals in the USA were delineated during the 1970s and 1980s (Auletta et al. 1993; Waters and Auletta 1981) and for food additives in 1982 (US FDA 1982). Classically, the first batteries included the following (1) a bacterial test for gene mutation, (2) either an in vitro test for chromosomal aberrations (based on the knowledge that ionising radiation and radiomimetic chemicals produced high rates of chromosomal aberrations even when induced mutation rates were relatively low) or a mammalian cell mutagenesis test and (3) a general test for DNA damage (US FDA 1982). An in vivo test was generally encouraged, with preference for a test for bone marrow chromosomal aberrations or micronucleus induction, based on the knowledge of a few chemicals that were uniquely active in vivo (ICH 1997; Tweats et al. 2007a). Much research effort was focused on development of appropriate mutagenicity testing methods that would detect a broad array of mutagenic chemicals. The classical series initiated by Hollaender, 'Chemical Mutagens: Principles and Methods for their Detection', was devoted to summarising these methodologies (Hollaender 1971).

By the time of the 1993 draft revision of the US Food and Drug Administration's (FDA) guidance on testing requirements for food and colour additives, the US FDA-recommended 'core' testing battery consisted of the following (1) a test for gene mutation in bacteria (*S. typhimurium*), (2) a test for gene mutation in mammalian cells in vitro, with the recommendation that the endpoint be based on an autosomal locus (so that events related to chromosomal interchanges could be detected) and (3) an assay for cytogenetic damage in vivo, with preference for a rodent bone marrow assay (US FDA 1993). By the year 2000, these so-called Red Book guidelines were finalised (US FDA 2000). Recently, Zeiger (2010) published a historical perspective on the development of genetic toxicology testing batteries in the USA, focusing on the policies of the Environmental Protection Agency (EPA). At the same time, the European, Japanese and Canadian recommendations were similar.

However, there were distinct differences in requirements both among regions and within different regulatory agencies within each region (DOH 1991; Shelby and Sofuni 1991). For example, the European recommendations generally included both an assay for gene mutation and an assay for chromosomal aberrations in mammalian cells (Kirkland 1993), while the Japanese relied on an in vitro mammalian cell chromosomal aberration assay and did not necessarily include the in vitro mammalian cell mutagenesis assay (Shirasu 1988).

Test practices regarding potential genotoxicity of pharmaceuticals, including test quality and assessment issues, have been delineated in a series of publications communicated by members of the German regulatory authorities (Madle et al. 1987; Müller and Kasper 2000; Müller et al. 1991). The evaluation spans the period between 1982 and 1997 and addresses nearly 600 new pharmaceutical entities. These publications summarise changes in test selection, improvements in test quality and shifts in the focus of test interpretation and assessment. The initial review (Madle et al. 1987), as well as its update (Müller et al. 1991), focused on deficiencies in test quality which was at that time considered to be a major issue. By the 1990s, this was no longer considered a major issue as it had been addressed by the International Workshop on Genotoxicity Testing and by revision of the OECD guidelines (see Sect. 6.7; OECD 1997). In addition, some genotoxicity systems which played a considerable role in the 1970s and 1980s, such as assays using yeast as indicator organisms, host-mediated assays, mutation tests using the fruit fly *Drosophila melanogaster*, sister chromatid exchange (SCE) tests in vitro or in vivo, chromosomal aberration analysis in bone marrow or spermatogonia, and dominant lethal assays, were little used by the 1990s. In part this reflects changes in test philosophy including a move away from assays involving cells of the germ line (Müller and Kasper 2000). This move was underpinned by evidence that germ cell mutagens appear to be a subset of somatic cell mutagens, so that compounds negative in somatic cell assays are very unlikely to be germ cell mutagens (Holden 1982; Shelby 1996; Waters et al. 1994).

In the early 1990s, the first ICH S2 guidances for genotoxicity testing were developed (see below), which achieved international agreement on a standard test battery and some aspects of protocol design (ICH 1995a, 1997). Meanwhile, experience with the test systems continued to accumulate. An evaluation addressing the experience in Germany with the review of tests for 776 new chemicals reviewed between 1982 and 1997 (Broschinski et al. 1998) focused on the frequencies of positive results in various standard in vitro systems. These data were compared with chemical structure characteristics and genotoxicity effects versus cytotoxicity. A later review of pharmaceuticals on the US market appeared to indicate a similar tendency to high frequencies of positive results, including the fact that ~20–30% of marketed pharmaceuticals seem to possess some kind of genotoxic potential especially in mammalian cells under in vitro conditions (Snyder and Green 2001). Since the accumulated knowledge indicates indirect means of genotoxicity in vitro, which may lack relevance in vivo, further evaluations have focused in a broader context on this issue. Kirkland et al. (2005) published an updated comparison of in vitro genotoxicity assay results with the ‘Carcinogenicity Potency Database’, the most comprehensive carcinogenicity database available (CPDB 2007).

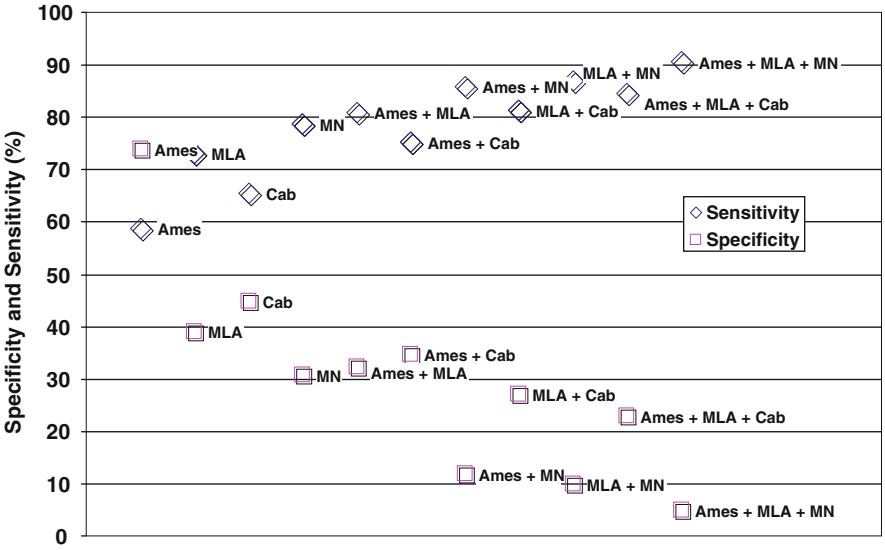


Fig. 6.1 Correlation data for the sensitivity and specificity for single in vitro genotoxicity assays and assay combinations (Ames *Salmonella* reverse mutation test introduced by B. Ames; MLA tk assay using the L5178Y mouse lymphoma cell line; Cab chromosome aberration test with different mammalian cell types; MN in vitro micronucleus assay with different mammalian cell types) (data according to Kirkland et al. 2005)

This evaluation showed in general terms that a battery of in vitro tests for genotoxicity can be pushed to high levels of sensitivity for detection of rodent carcinogens (sensitivity), but this sensitivity came at the price of inappropriately increasing the likelihood of obtaining a positive genotoxicity result for noncarcinogens (specificity) (Fig. 6.1). Matthews et al. (2006a, b) confirmed these results using a database that included proprietary data on pharmaceuticals. If the test approach is to maximise sensitivity for hazard identification, there is a heavy emphasis on the need for a weight-of-evidence approach for risk assessment (Bergmann et al. 1996; Cimino 2006; Dearfield and Moore 2005; US FDA 2006; Kasper et al. 2007; Müller et al. 2003). In the views of many scientists, there is a growing lack of confidence in the results that come out of regulatory in vitro genotoxicity tests and a need to refine the conditions for in vitro mammalian cell genotoxicity tests to optimise their predictive value or to introduce new assays (Kasper et al. 2007; Kirkland et al. 2005, 2006, 2007a; MacGregor et al. 2000; Müller et al. 2003). In part, this is driven by the honest desire to minimise use of animals in testing. In Europe, efforts in this context culminated in a publication from a workshop held under auspices of the European Centre for Validation of Alternative Methods (ECVAM) entitled ‘How to reduce false positive results when undertaking in vitro genotoxicity testing and thus avoid unnecessary follow-up animal tests: Report of an ECVAM Workshop’ (Kirkland et al. 2007a). In addition, the International Life Sciences Institute (ILSI http://www.ilsi.org/Pages/AboutILSI.aspx) has instituted a working group that tackles ‘Relevance and follow-up of positive results in in vitro genetic toxicity assays: an ILSI-HESI initiative’ (Thybaud et al. 2007a, b).

There are two major sets of internationally applicable regulatory guidelines for genotoxicity testing and two major international scientific processes that dominate the regulatory landscape:

1. The Organisation for Economic Co-operation and Development (OECD) test guidelines
2. The International Conference of Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) guideline(s) for genotoxicity testing of pharmaceuticals
3. The International Workshop on Genotoxicity Testing (IWGT)
4. The International Programme on Chemical Safety (IPCS) under auspices of the World Health Organization (WHO) (Ashby et al. 1996)

These guidelines and processes are selected because they have been driving the scientific process (IWGT and IPCS) and have set internationally acknowledged standards of testing that go beyond country or regional borders (ICH and OECD). Other institutions such as the Health and Environmental Safety Institute (HESI) under the umbrella of the International Life Sciences Institutes (ILSI http://www.ilsi.org/Pages/AboutILSI.aspx) also provide means of internationally acknowledged and coordinated efforts to improve the regulatory sciences of mutation research. This chapter does not focus on genotoxicity guidelines for compounds for other purposes such as pesticides, new chemicals and food additives. The reader is referred to other review articles that cover these regulations in more detail (e.g. Cimino 2006; Zeiger 2010) or to the specific guidelines such as those from the United Kingdom's Committee on Mutagens (UKCOM 2000), the US Environmental Protection Agency's (EPA's) Office of Prevention, Pesticides and Toxic Substances' health effects' test guidelines series 870 (EPA 2008) and the European Union's (EU) guidelines for the testing of chemicals under Annex V, Part B (EC 2008). It might be worthwhile to mention here that regulatory guidance documents on genotoxicity are partly influenced by general considerations prevailing in the society such as the three R's (reduction, replacement and refinement) to optimise use of animals in safety testing and to reduce their burden. For example, this has led in the European Union to a ban of animal testing for cosmetics, which require a focus on in vitro genotoxicity testing only. This approach will require work on more predictive in vitro tests for the purpose of risk assessment (Kirkland et al. 2008).

6.1.2 Genotoxicity Guidelines on Pharmaceutical Testing and the ICH Process: Pre-ICH Position and Background for the Need for Harmonization of Genotoxicity Guidelines

In the three regions covered by the ICH exercise, that is, Europe, Japan and the USA, different approaches had evolved with regard to the genotoxicity tests required for the registration of new pharmaceuticals for human use, and differences

were present in many aspects of the protocols used for individual tests. For instance, a small survey of European pharmaceutical companies in 1991 identified over 60 differences in aspects of genotoxicity testing between the three regions from their experience with national regulators. By 1992, the year of the first meeting of the ICH genotoxicity expert working group (EWG), these differences had led to major problems for the pharmaceutical industry, as the market had become more globalised, and the same dossier to support registration may have been acceptable in one region but rejected in another, as local requirements had not been met. This had led to redundant testing and a loss in time and effort in bringing new medicines to the market.

In Europe, several countries had issued national guidelines, for example, the UK Department of Health and Social Security (DHSS) had published 'Guidelines for Testing of Chemicals for Mutagenicity' in 1981 (DHSS 1981), with a revised version published in 1989 (DH 1989). In 1981 and 1987, the European Commission issued a similar guideline applicable to medicinal products (EEC 1987). In Japan, genotoxicity guidelines had been existence for a number of years, such as the Japanese Ministry of Health and Welfare guidelines (JMHW 1984). In the USA, there were no specific guidelines for genotoxicity tests for pharmaceuticals, although the US FDA Center for Food Safety and Applied Nutrition (CFSAN) had guidance for genotoxicity testing of foods and food additives in the so-called Red Book (US FDA 1982, 1993), and other US agencies such as the Environmental Protection Agency also had issued guidelines on genotoxicity testing (USEPA 1982a, b). Although there were many common features among these guidelines, there were also differences, and there was no attempt to look for harmonization at this level before the ICH process began. The guidelines that had been produced were constructed on the basis of academic advice available in the relevant regions and thus reflected the expertise and experience available locally. Pharmaceutical companies found that without formal guidelines in the USA, there was often a difference in views between different reviewers within the FDA, so what might have been acceptable in one Division for a previous submission may not necessarily be acceptable for the next submission to another Division. In Japan, submissions would be checked for consistency with Japanese guidelines with no flexibility for a slightly different approach based on the characteristics of a particular chemical class. In addition, it was difficult to have access to and discussions with the genotoxicity experts within the agency, on specific submissions. It was informative to the ICH EWG to understand that 'guideline' translated to 'rule' in Japanese rather than a 'guidance' that would permit some flexibility on a case-by-case basis. In Europe, there were preferences for some approaches that did not meet with the needs of the other agencies.

Separate from the Pharmaceutical Regulatory agencies, as mentioned previously, there were several excellent international initiatives to harmonise genotoxicity protocols, such as the OECD guideline series from 1983 (OECD 1983, 1984, 1986) which, by the time the ICH exercise started, had issued guidelines on most of the established genotoxicity assays, following a template approach to ensure that key aspects had been covered including the analysis of the data produced. Although these guidelines were not strictly intended for pharmaceuticals, they did provide

useful information for the ICH EWG. Another initiative that began during the time the EWG was active and which contributed to the deliberations of the EWG was the International Workshop on Genotoxicity Testing (IWGT), as discussed earlier. With the emphasis on test procedures or protocols, this initiative brought together scientists from academia, industry and from regulatory authorities to deliberate on genotoxicity test issues and seek consensus on solutions (Kirkland et al. 2007b, c, 2011). Apart from providing views on key aspects of genotoxicity test protocols, this initiative helped to encourage and foster joint working of key opinion leaders in the field and provide extra impetus for collaborative harmonization efforts.

The United Kingdom Environmental Mutagen Society (UKEMS) had also issued a series of influential guideline books in the early 1980s that provided details of genotoxicity testing, including the quality of the materials tested, the test protocol and data processing and presentation (UKEMS 1990, 1993).

Most of the existing regulatory guidelines for test batteries were built on the premise that the established in vitro tests can identify those drug candidates that possess genotoxic potential. These in vitro tests would include the *Salmonella*/Ames test and one or more mammalian cell assays. Both types of test were seen as needed, as there are some genotoxins that interact with components in mammalian cells that are absent in bacteria. All guidelines also included in vivo tests to show whether this potential is realised in the whole animal. The differences in approach came from what cells to use, what mammalian tests were vital for a core test battery and whether information was comparable between test systems. There was little guidance on what to do to follow up on positive results obtained in vitro, although it was accepted that in vivo tests in one tissue such as the bone marrow may generate false negatives, as the test compound or its metabolites may never reach that tissue to express an effect, so further in vivo testing may be needed.

The ICH EWG was composed of six representatives, one each from the pharmaceutical industry associations in the three regions and one each from the regulatory authorities from the three regions. Each of these representatives was backed by a committee of experts in their constituency that identified issues for harmonization. Following the first meeting held in Tokyo in December 1992, a total of 11 major topics were identified for harmonization in the three regions. In order to address these topics, the EWG agreed to use the best available science and, if necessary to achieve agreement, would commission new studies or new analyses of data to address specific issues.

6.2 The ICH S2 A Guidance (1995)

As mentioned in the previous section, the ICH guideline process for S2 was started in 1992. By 1995, several years of negotiation had passed. Yet the time was not sufficient to achieve agreement between the three regions involved over one major item, the standard battery of genotoxicity tests for pharmaceuticals. In 1995 it was not clear whether there would be a chance to achieve agreement over this major item

within a foreseeable time frame. However, all parties agreed that there were valuable agreements over many specific aspects of regulatory genotoxicity testing of pharmaceuticals at that time and that these achievements should not be lost. To this end, the ICH Steering Committee agreed in 1995 to split the ICH guidelines on genotoxicity and to proceed with different pace towards two final guidelines. Hence, an ICH guidance on 'Specific Aspects of Regulatory Genotoxicity Testing of Pharmaceuticals', called in short ICH S2A, was finalised at Step 4 in 1995 (ICH 1995a). Within the same year, this guidance came into operation in all three regions. Because ICH S2A addressed at the time only specific aspects of existing guidances, the introductory paragraph in this first ICH genotoxicity guidance still mentions the old regulatory requirements in the EU and Japan and procedures followed by the US FDA.

The S2A guideline addressed both protocol issues and strategy issues. The guidance given is summarised below.

6.2.1 Protocol Issues

6.2.1.1 Target Tissue Exposure

In pre-ICH guidelines, the only commonly used monitor of exposure of the target tissue, for example, the bone marrow, had been the demonstration of a biological effect in that tissue, such as a reduction in the formation of polychromatic (immature) erythrocytes. Often such effects cannot be observed, even at near lethal doses of a test compound due to limiting toxicities in other tissues. Toxicokinetic measurements of drug-related compounds, such as blood or plasma levels (Probst 1994) or direct measurement of these in target tissues or from autoradiographic assessment of the target tissue following exposure to radio-labelled drugs, can show that significant exposure has occurred. This can be in the absence of a change in a local biological marker. It was agreed that where the compound in question is negative in the in vitro part of the test battery, exposure can be inferred from standard rodent ADME tests. For compounds where some component of the in vitro test battery has shown positive results, it was agreed that test-specific toxicokinetic data should be generated. If adequate exposure in the target tissue cannot be achieved due to poor bioavailability or extensive protein binding, in vivo tests may have little value.

6.2.1.2 Testing into the Insoluble Range

6.2.1.3 Target Concentrations and Desired Levels of Cytotoxicity for In Vitro Tests

For freely soluble, non-toxic compounds, the desired upper treatment levels were defined as 5 mg/plate for bacteria and 5 mg/ml or 10 mM (whichever is the lower) for mammalian cells. These concentrations were based on legacy evidence that cer-

tain genotoxic carcinogens were not detected until these upper limits were tested. Higher levels were not deemed necessary as there is a risk of adverse effects on osmolality or other changes that could generate false results (see later).

It was recognised by the EWG that at very high levels of cytotoxicity in vitro, events other than genotoxicity can interfere with the proper evaluation of the relevant genetic endpoints. These include those associated with apoptosis and endonuclease release from lysosomes, which can induce chromosome damage. Various recommendations were made as follows: In bacterial mutation assays, the highest concentration of the test compound is desired to show evidence of toxicity. This can be manifested as a reduction in the number of revertant colonies and/or a clearing or diminution of the bacterial background lawn. For in vitro cytogenetic tests using established cell lines, the top tested concentration should induce a greater than 50% reduction in cell number or culture confluency. For tests using human lymphocytes, an inhibition of the mitotic index by greater than 50% is considered sufficient. In mammalian cell mutation studies, the recommended highest concentration should produce at least 80% toxicity. A note of caution was added to the guidelines with interpreting data at levels of survival lower than 10%, where misleading results can be obtained. There was a discussion of the methods of obtaining toxicity information, but cloning efficiency and calculation of relative total growth were deemed the most reliable measurements. These levels of toxicity were based on legacy evidence that some genotoxic carcinogens are not detected until such concentrations are tested.

6.2.1.4 Acceptable In Vivo Tests for the Detection of Clastogens

Prior to ICH there were some regulatory disagreements concerning the comparability of in vivo micronucleus tests with tests using metaphase analysis. It was agreed that regulators can accept the validity of either test for measuring chromosome damage. In addition, measuring micronucleated polychromatic erythrocytes in peripheral blood was accepted as a viable alternative to bone marrow testing in the mouse, as the mouse spleen is incapable of efficiently removing such cells from the circulation.

Data on over 100 compounds, again mainly from Japanese laboratories, had shown that under normal circumstances male rodents are sufficient for use in the bone marrow micronucleus test (CSGMT 1986). Although quantitative differences between the genders were seen occasionally (usually males were more sensitive than females), qualitative differences were rare. These differences are associated with differences in toxicokinetics or metabolism. Thus, it was agreed that if such differences are seen in toxicology testing, both sexes should be employed, otherwise males are sufficient. If gender-specific medicines are to be tested, the appropriate sex should be used. It was hoped that these clarifications would enable some saving in animal resources.

Most micronucleus test data have been generated in the mouse, but at the time the S2A guideline was being constructed, there was deemed to be sufficient evidence from micronucleus studies in the rat to allow either species to be used. The use of the rat was seen as beneficial, as this is the species of choice for most rodent toxicology testing, and thus, there was useful ADME data generated routinely that

would also aid the interpretation and validation of the micronucleus assays. It was recognised that there are a small number of species-specific genotoxins (e.g. Albanese et al. 1988). Thus, this was an area where the EWG thought that if more data of this type accumulates in the future, the decision of whether to use one or more species for micronucleus testing in vivo may need to be revisited.

6.2.2 Strategy Issues

6.2.2.1 The Base Set of Strains Used in Bacterial Mutation Assays

It was agreed that the standard, acceptable strains for use in bacterial mutation assays should be able to detect both base change and frameshift point mutations. The target genes for mutation in the commonly used *Salmonella* strains such as TA1535/TA100 (detect base changes) and TA1537/TA98 (detect frameshift changes) involve mutations at G–C base pairs. There is good evidence that some mutagens causing oxidative damage preferentially mutate A–T base pairs (e.g. Wilcox et al. 1990). Thus, it was recommended that *Salmonella typhimurium* TA102 or *E. coli* WP2 *trpE* strains, all of which have target genes with A–T mutations, should be included in the standard set of strains for these assays. Analysis of a database of 5,526 compounds held by the Japanese Ministry of Labour and shared with the EWG had shown that 7.5% of the bacterial mutagens identified mutated *E. coli* WP2 *uvrA*, but not the 4 standard *Salmonella* strains. The compounds in question were deemed to be likely genotoxins of concern.

6.2.2.2 Timing of Genotoxicity Testing

The genotoxicity EWG liaised with the EWG for the M3 guideline, which defined the nonclinical safety studies for the conduct of human clinical trials for pharmaceuticals. It was agreed that prior to first human exposure, in vitro tests for the evaluations of mutations and chromosomal damage are generally needed. If an equivocal or positive finding occurs, additional testing should be performed. The standard battery of tests for genotoxicity should be completed prior to the initiation of phase II studies.

6.2.2.3 Guidance on the Evaluation of In Vitro Test Results

The EWG was aware that all genotoxicity assays generate false-positive and false-negative results. False-negative results can occur in vitro due to inadequacies of the in vitro metabolic activation system; the absence of the target for genotoxicity, for example, the cell division apparatus found in mammalian cells but not bacterial cells; etc. The test battery approach is designed to overcome these problems. Even at the time of the construction

of the S2A guideline, it was known that there can be an excess of false positives in in vitro mammalian cell tests for several known reasons. Thus, the S2A guideline went to some lengths to define conditions, to define the pitfalls and provide guidance on the interpretation of these studies. Various aids to interpretation were listed including:

1. Is the increase in response over the negative or solvent control regarded as a meaningful genotoxic effect?
2. Is the response concentration related?
3. For weak/equivocal responses, is the effect reproducible?
4. Is the positive result a consequence of an in vitro specific metabolic activation pathway?
5. Can the effect be attributed to extreme culture conditions that do not occur in in vivo situations, for example, extremes of pH; osmolality; etc.?
6. For mammalian cells, is the effect only seen at extremely low survival levels?
7. Is the positive result attributable to a contaminant? (This may be the case if the compound shows no structural alerts or is weakly mutagenic or mutagenic only at very high concentrations.)
8. Do the results obtained for a given genetic endpoint conform to that for other compounds of the same chemical class?

The point of this guidance was to show that increases in mutants or in the amounts of chromosomal damage seen in these in vitro assays used do not always signify that the compound in question is genotoxic. The results always need to be seen in context and evaluated for biological relevance.

Further guidance is given for negative results in vitro as follows (1) Does the structure or known metabolism of a compound indicate that standard techniques may be inadequate and that modifications of the S9 mix may be required for detection, for example, the use of hamster S9 for screening diazo compounds? (2) Does the structure or known reactivity of the compound indicate that the use of other protocols may be appropriate?

6.2.2.4 Guidance on the Evaluation of In Vivo Test Results

The EWG recognised that the number of validated in vivo tests available at the time was limited, in particular, there were no validated in vivo tests for gene mutation, although mutation assays using transgenes were at various stages of development.

However, if a compound had been tested in vitro with negative results, then a bone marrow or peripheral blood cytogenetic assay would be sufficient. If, however, some element of the in vitro tests had given a biologically significant response, further in vivo tests were required. This was in recognition of the less-than-optimum sensitivity of the cytogenetic assays, where exposure may not be optimum. Thus, this additional testing should be in tissues other than the bone marrow. Examples that might be considered included the ex vivo liver UDS test, DNA strand-break assays and ^{32}P-post-labelling studies. If in vivo and in vitro tests do not agree, then the differences should be considered on a case-by-case basis.

The guidance given stated that the assessment of genotoxic potential of a compound should take into account the totality of findings and acknowledge the intrinsic values and limitations of both in vitro and in vivo tests. Thus, there were several issues that were outstanding after the issue of the S2A guideline. It took two more years of negotiation and some new experimental data to allow consensus to be reached on defining the standard battery of tests and some aspects on elements of the protocols for the selected in vitro tests.

6.3 The ICH S2B Guidance (1997)

While the ICH S2A guidance was already in use, the ICH expert working group on genotoxicity continued to work on harmonization of the standard battery of genotoxicity tests for pharmaceuticals. As mentioned above, no in vitro cell mutation assay was required in Japan, and therefore, Japanese laboratories had no experience with such assays to be included as a member of the standard battery. To overcome this situation, Japanese National Institute of Hygienic Sciences (National Institute of Health) and JPMA started to evaluate the MLA using L5178Y cells, especially for an evaluation of using them as an alternative to the in vitro chromosomal aberration assay. A total of 40 compounds were selected for a collaborative study, among which were 33 that were positive in in vitro chromosome aberrations studies, but negative in bacterial mutation tests. Out of the 33 chemicals, 19 were positive in the MLA, 6 were equivocal and 8 were negative. Some of the negative compounds were positive only in the extended sampling time of 24 h in the chromosome aberration test, and some of them were nucleoside analogues or polyploidy inducers that may have targets in the cell other than DNA and also require extended exposure times to exert their effects. Therefore, a repeat MLA set of studies was carried out in which exposure was extended from the normal 3–4 h to 24 h (in the absence of rat liver S9, as the cells find S9 toxic over long incubation periods) with a subset of chemicals that fell into this category. Ten of these compounds were positive at 24 h in the MLA. This outcome showed that the MLA could be used as an assay for detecting clastogens, equal to the in vitro mammalian cell chromosomal aberration test, as long as the MLA protocol contained a 24 h sampling time in addition to the normal 3–4 h sampling time (Honma et al. 1999). Further negotiations, collaborative studies and changes in perceptions enabled the EWG to finalise the second guidance on 'A Standard Battery for Genotoxicity Testing of Pharmaceuticals' (ICH S2B) by their 17th meeting in July 1997, just in time for the ICH4 conference in Brussels.

6.3.1 The Standard Battery of Tests

The standard 3-test battery agreed upon by the EWG is shown in the list below. Thus, two in vitro tests are required, one for the detection of bacterial mutation (as defined in ICH guideline S2A) and one for chromosome damage induction in

mammalian cells, which could be either an in vitro chromosome aberration assay or the in vitro mouse lymphoma tk assay. Such assays are required for those genotoxins that target aspects of the cellular machinery, not present in bacterial cells. The battery is completed by an in vivo test for chromosome damage using rodent haemopoietic cells. An in vivo test provides additional relevant factors that are difficult to model in vitro, such as absorption, distribution, metabolism and excretion, which may influence the genotoxicity of a test compound. This test can be either an analysis of chromosomal aberrations in bone marrow cells or an analysis of micronuclei in bone marrow immature erythrocytes or in peripheral blood erythrocytes. The EWG recognised that there were new tests under development, in particular the in vitro micronucleus assay, that could be included in the future as an alternative chromosome damage (and aneuploidy) detecting test.

The Standard 3-Test Battery

1. A test for gene mutation in bacteria
2. An in vitro test with cytogenetic evaluation of chromosomal damage with mammalian cells *or* an in vitro mouse lymphoma tk assay
3. An in vivo test for chromosomal damage using rodent haematopoietic cells

For compounds giving negative results on the completion of this 3-test battery, this was regarded as sufficient to demonstrate the absence of genotoxic activity. However, the EWG recognised that there are sets of compounds that may require a different test battery. Five conditions were identified needing a modified approach as shown in the following list:

Situations Where Variations in the Standard Battery May Be Required

1. Limitations in the use of bacteria
2. Compounds bearing structural alerts for genotoxicity that are negative in the standard test battery
3. Limitations in the use of standard in vivo tests
4. Additional genotoxicity testing to determine the mode of action of carcinogenicity
5. Structurally ‘unique’ chemical classes

For the first of these, when testing compounds that are excessively toxic to bacteria, such as antibiotics, or when testing compounds that target specific components of mammalian cells that have no equivalents in bacterial cells, bacterially based assays have limited value. Thus, for such compounds, the EWG recommended that both types of mammalian cell assays should be completed, that is, a test for gene mutation and a test for chromosome damage. However, as some genotoxins still induce detectable mutations in bacterial cells at very low concentrations, for example, nitrofurans, a bacterial mutation test should still be completed.

For compounds that possess known structural alerts for genotoxicity and give unexpected negative results in the standard battery, limited further testing is recommended. This may include appropriate protocol modifications such as altering the

components or composition of the S9 mix employed or based on any other chemical class-specific information.

If a test compound is known to be very poorly absorbed systemically such that the standard in vivo assays are inappropriate because there will be little or no exposure to the target cells, then an all in vitro test battery was recommended, again including both mutation and chromosome-damaging mammalian tests.

In some cases, compounds elicit carcinogenicity in the lifetime studies in rats and/or mice, where the mode of action is likely to be through genotoxicity. In rare cases, such compounds may have given negative results in the standard battery. It was agreed that further genotoxicity studies may be helpful to elucidate the mode of action of carcinogenicity. Such studies may involve modified in vitro testing; in vivo genotoxicity tests in the target organ for carcinogenicity, for example, the ex vivo liver UDS test; ^{32}P-post-labelling studies to detect DNA adducts; or mutations in transgenes.

On rare occasions, a completely novel compound in a unique chemical class will be submitted for assessment. For such compounds, the standard test battery may not give full confidence that such compounds are devoid of genotoxicity, and additional testing may be warranted.

6.3.2 Standard Procedures for In Vitro Tests and Confirmatory Testing

The S2B guideline also gave guidance on standard procedures for in vitro tests, including the issue of confirmatory testing. It is held as good scientific practice that, irrespective of the initial test result, the outcome of scientific experiments should be verified by conducting the same test again, employing the same or modified procedures. However, the EWG reasoned that the state of the art of the routinely employed assay systems is such that the protocols are well defined, standardised, understood and are well controlled. Thus, whereas equivocal results should always be explored by repeat testing, with possible modifications of the protocol such as modifying the conditions for metabolic activation and changing the spacing of the test concentrations, there is less of a need to repeat clearly positive or clearly negative results. For positive results, it was left to the discretion of the testing sponsor whether or not to confirm the result. For negative results, it was suggested that expanded use of the range-finding test for mutation assays could provide the necessary element of repetition. Thus, as well as using the range-finding assay to inform on the selection of concentrations to test, there could also be an analysis of induced mutations to add to the assessment of mutagenesis in the main assay. If both are negative, this was deemed sufficient confirmation. For in vitro cytogenetic assays, it was recognised that the required protocol has acceptable built-in confirmatory elements, for example, sampling at different exposure times. It was stressed that an acceptable mouse lymphoma assay should include colony sizing to distinguish large colony mutants,

which arise predominantly through point mutation, from small colony mutants that arise predominantly from chromosome damage.

Since this time, it is understood that both S2A and S2B ICH guidelines should be read and applied together. Further, since 1997, these two ICH guidelines have fully replaced previously existing genotoxicity guidelines for pharmaceuticals in the three ICH regions.

6.4 Experiences with the Use of ICH S2A and S2B Guidances in the ICH Regions

The introduction of the standardised battery and test protocols in the ICH S2A and S2B and 1997 OECD guidelines was a great step forward in enabling international acceptance of genotoxicity test packages for sponsors. Two main issues that continue to cause difficulties are the persistent high frequency of positive results in the in vitro assays for chromosome breakage and L5178Y (mouse lymphoma cell) *tk* mutation (MLA) and variability between and within regulatory agencies in the requirements for testing to follow up positive results, both in the timing and in the type and quantity of test data required to advance through clinical trials.

As noted above, one quarter of 352 marketed pharmaceuticals (excluding drugs expected to be genotoxic such as anticancer agents and nucleosides, as well as steroids and biologicals) in the 1999 US Physicians' Desk Reference (PDR) were positive in the in vitro chromosome aberration assay or the MLA (Snyder and Green 2001). The lack of predictive value and poor specificity for rodent carcinogenicity of these assays has been discussed above (see Kirkland et al. 2005; Matthews et al. 2006a, b). Although more standardised test protocols had been put in place after the 1992 IWGTP workshops and during the finalisation of the OECD guidelines (1997) and of the ICH S2A and S2B guidances (ICH 1995a, b, 1997), this frequency apparently did not change. An analysis provided by P. Kasper of BfArM of data submitted to the German regulatory agency showed that for 596 compounds from 1995 to 2005, 181 compounds were positive in at least one in vitro clastogenicity assay (chromosome aberrations or MLA), that is, 30% of the drugs. Further, an analysis of drugs in the PDR was extended to more recent submissions, and it was shown in 2005–2006 that there was no change in the conclusions. The updated analysis (PDR information from 1999 through 2008) for 545 marketed drugs again reported that 19–26% were positive in in vitro mammalian cell assays (Snyder 2009). Japanese regulatory scientists reported that they had a similar experience.

For pharmaceutical companies, this meant that in some cases promising drugs were dropped from development or considerable amounts of follow-up testing, including animal tests, were required to advance the compound. The ICHS2 guidances provided factors to consider in assessing the significance of such positive results, but there was a lack of consistency in regulatory approaches to

the mammalian cell assay data. In some cases, sponsors were asked to perform transgenic mouse carcinogenicity assays before carrying out repeat-dose clinical trials, causing extensive delays (Jacobson-Kram and Jacobs 2005), although results routinely proved negative in p53+/− mouse carcinogenicity studies (Jacobs 2005). The FDA issued an internal guidance on 'integration' of genetic toxicology results, providing advice to its reviewers on how to assess weight of evidence or approach follow-up testing to understand mechanism or mode of action without having to go as far as doing cell transformation assays or carcinogenicity assays (Jacobson-Kram and Jacobs 2005; US FDA 2006), but variable regulatory requirements continued.

It has long been clear that the in vitro mammalian cell assays are responding to events that do not represent direct interaction with DNA. Although altered pH and osmolality are often highlighted, these are not likely to be involved because the phenomena were reported in the 1980s and formal recommendations for controlling pH and avoiding increased osmolality (10 mM limit) have been in place since the early 1990s. In many cases the mode of action appears to involve indirect effects, such as inhibition of DNA synthesis, and non-specific effects related to toxicity and energy metabolism. It has been postulated that non-covalent binding (intercalation) and potential topoisomerase inhibition play a role in chromosome damage by chemicals that do not directly damage DNA (Snyder 2010). These indirect effects are likely to have a threshold below which the mechanisms are not operative and are generally unlikely to occur under pharmacologically relevant conditions in vivo. The high frequency of positive results with compounds that are known not to damage DNA and the lack of correlation with rodent carcinogenicity data have led to a lack of confidence in the in vitro mammalian cell assays.

The kinds of efforts that had preceded the introduction of the ICH S2 guidances continue, under the auspices of IWGT, ECVAM, ILSI–HESI and other groups, to improve protocols and test strategies by collecting data and assessing experience with existing assays, with follow-up strategies and with development of newer assays (e.g. Kirkland et al. 2007a; Thybaud et al. 2007a, b). To make use of another 10–15 years of experience, a renewed effort began to update the ICH S2 guidance.

6.5 ICH S2(R1): Reasoning for Revision of the ICH S2A and B Guidances

Within 10 years of the ICH guidelines for genotoxicity testing being in operation, it was realised that advances in genotoxicity testing and interpretation would require a maintenance process. This process was initiated in 1999 and seriously at the end of 2006 and has resulted in a new single draft ICH 'Guidance on Genotoxicity Testing and Data Interpretation for Pharmaceuticals Intended for Human Use' (ICH 2008). For the interested reader, the original statement of the issues requiring a revision process of the ICH S2 guidelines, which were presented to the ICH Steering Committee in the summer of 2006 are reproduced below.

6.5.1 ICH Concept Paper for Revision of the ICH S2A and S2B Guidances (2006)

6.5.1.1 Statement of the Perceived Problem/Issues To Be Resolved

Genetic toxicity testing relies largely on short-term tests, thus new technical knowledge tends to develop rapidly. In addition, scientific understanding of the nature and relevance of different types of genetic damage and different modes of action involved in the process of mutagenesis is also improving. The ICH guidances concerning genotoxicity were finalised in 1995 (S2A) and 1997 (S2B). Since then there have been new developments and a wealth of data on both in vitro and in vivo genotoxicity assays with the potential to add value to the guidance given in the original guidelines. These include the in vitro micronucleus test for the detection of genotoxic compounds (clastogens and aneugens) and assays that are applicable to a variety of tissues in vivo, i.e. the comet assay for DNA strand breakage, and transgenic mutation models.

The in vitro mammalian cell tests recommended in the S2B guidance are not fully capable of detecting aneugens. Thus, the in vitro micronucleus test may provide an option that facilitates the detection of this important class of genotoxins better than with the existing models. The preferred in vivo tests described in the S2B guideline measure chromosome damage in the bone marrow, and for follow-up testing of in vitro positive compounds, DNA repair in the liver as these were the only validated models at the time. Already then, it was highlighted that these tests may not reflect some tissue-specific genotoxins. The capability of the new in vivo tests to be applied to the tissue of choice (or high exposure) such as the GI tract for 'site of contact' genotoxins in case of oral administration, will provide a better assessment of genotoxic potential in vivo.

Another severe problem, which emerged during the last years in regulatory testing, is the high rate of positive findings especially in the in vitro mammalian cell tests recommended in the S2B guidance, i.e. the mouse lymphoma test and the chromosomal aberration test. The interpretation of the relevance of many of these in vitro findings has been frequently debated and extensive in vivo and/or mechanistic follow-up studies are required. Several recent reviews confirm oversensitivity and lack of specificity of both in vitro test models (Kirkland et al. 2005; Matthews et al. 2006a, b). A more rational approach to testing conditions and of interpretation of the genotoxicity data is required either by application of new techniques and/or modification of existing models/approaches or by deleting the requirement for such testing.

The purpose of the ICHS2 A and B revision is to achieve several goals. First, it should reduce the numbers of animals used in routine testing by improving the current procedures (limitation in the number of animals used as positive controls) and clarifying the follow-up testing in case of positive findings. Second, it should avoid or more adequately manage/interpret the irrelevant findings in order to reduce barriers in early drug development by improving risk assessment for carcinogenic

effects that have their basis in changes in the genetic material. Finally, it should update and improve internationally agreed upon standards for follow-up testing and interpretation of positive results, especially from in vitro assays, in the standard genetic toxicology battery.

6.5.2 Guiding Principles for a New ICH S2(R1) Guideline

Several underlying principles were agreed by the expert working group as a basis for revising the ICH S2A and S2B genotoxicity guidelines for human pharmaceuticals and to generate a single, revised ICH S2(R1) guidance. These were built on good tradition-based principles used in the ICH and IWGT processes:

1. In designing a test battery, it is important to consider prediction of compounds of concern for human health. In the past, tests or protocols have been selected for their ability to predict rodent carcinogenesis. We do not have a good replacement for rodent carcinogenicity data yet, but we can take into account information that helps determine whether a given carcinogen is acting primarily through genotoxic mechanism(s) and so may indicate a potential risk to humans, or non-genotoxic mechanism(s), more likely to have a threshold and ultimately be considered non-relevant to people. Short-term genotoxicity assays have no role in identifying such rodent carcinogens.
2. Affirming the statement in the original ICH guidance, the battery is primarily aimed at detecting potential genotoxic carcinogens, but in so doing will protect against potential germ cell mutagens. All the known germ cell mutagens are positive in bone marrow chromosome damage assays (Adler and Ashby 1989; Shelby 1996: Waters et al. 1994; Holden 1982).
3. In choosing tests or test conditions appropriate for a battery, we should consider the battery approach overall, that is, it is not essential that each individual test be able to detect all genotoxicants/compounds of concern for carcinogenicity, provided another test in the required battery effectively detects it.
4. We need to reduce our reliance on in vitro assays, especially mammalian cell assays, carried out under somewhat extreme conditions on the principle of hazard identification, and consider test/protocols that identify potential genotoxic effects under more realistic conditions that provide information more useful to human risk evaluation.
5. Exceptions should not drive the design of test batteries or protocols. For example, if certain tests or modified protocols are required to detect nucleoside analogues, these may be given as an example where specialised test strategies may be needed but not used to change the design of a strategy that effectively detects all other known classes of genotoxins.
6. It is appropriate that a battery continue to contain a bacterial gene mutation assay and a mammalian cell genotoxicity assay. As when the battery concept first developed, the justification for the mammalian cell assay is twofold: partly based on the observation of a few genotoxins that are more effectively detected in

mammalian cells and partly on the principle that since many pharmaceuticals are aimed at mammalian cell targets, genotoxicity specific to mammalian cells may well occur. Neither is dependent on whether the mammalian cell assay is performed in vitro or in vivo. Traditional batteries include a test for gene mutation in bacteria and a test for mammalian chromosome damage based on (a) the observation that the combination of the bacterial mutation test and a chromosome damage assay in vivo predicts the majority of known human carcinogens (reviewed in Shelby and Zeiger 1990; Tinwell and Ashby 1991) and correlates well with rodent carcinogens (Matthews et al. 2006a, b) and (b) the considerations that (1) a wide range of types of genetic events that occur in tumorigenesis and in heritable mutations are detectable in mammalian but not bacterial cells and (2) many pharmaceuticals have mammalian cell targets. In practice, only a few examples of compounds that are more effectively detected in mammalian cells exist and are not typically pharmaceuticals (e.g. metals), and numerous studies have demonstrated conclusively that adding more in vitro tests to the Ames test does not improve predictivity for rodent carcinogenesis (Matthews et al. 2006b).

7. Reaffirming the philosophy behind the original ICH battery, a mammalian cell gene mutation assay is not a required part of the test battery. The MLA was seen as an equally acceptable alternative to the in vitro chromosome breakage study. While some chemicals are more efficient at inducing mutants and others are more efficient at inducing larger scale chromosome damage, the case for endpoint specificity ('gene mutagens' vs. 'clastogens/aneugens') has been examined repeatedly without producing convincing evidence. For prediction of rodent carcinogenicity, the most accurate positive predictor remains the Ames bacterial mutagenicity assay. All of the known rodent germ cell mutagens are detectable in the rodent bone marrow cytogenetic (chromosome breakage or micronucleus) assays (see above).
8. While OECD guidelines were developed for genotoxicity assays in parallel with the ICH guidances in the early 1990s and there was a purposeful attempt to develop methods applicable to genotoxicity testing for all types of chemicals (industrial, agricultural, medical), it is clear that certain attributes of pharmaceutical testing justify specific modifications for drugs, and differences from existing OECD guidelines are pointed out and justified in the revised ICH guidance.

A text for a single new ICH S2(R1) guideline was finalised by the ICH EWG (core members Philip Bentley, Sheila Galloway, Jerry Frantz, Makoto Hayashi, Masamitsu Honma, David Jacobson-Kram, Peter Kasper, Lutz Müller, Timothy Robinson, Shigeki Sawada, Veronique Thybaud, Jan-Willem van der Laan, Akihiro Wakata) and published for comments in the regulatory arena of the EU, Japan and the USA in early 2008. Further to an internal issue resolving process within the US FDA, the guidance was finally adopted in the ICH process at the end of 2011 and is now becoming practice in all regions.

In the following, the main principles of the revised ICH S2(R1) guidance are laid down and discussed. It is understood that some of these principles will bring about

changes to genetic toxicology testing and its regulatory use, and we may see genetic toxicology guidance for purposes other than testing of pharmaceuticals follow these rationales.

1. The Ames test continues to be an elementary and indispensable part of regulatory testing. However, there is no continued need to repeat fully adequate negative Ames tests in an independent experiment.

While it is clear that there are some differences between mammalian cells and bacteria in regard to metabolism and DNA repair processes, it continues to be acknowledged that there is no suitable alternative for the bacterial reverse mutation test (Ames test). The Ames test is the most widely used test for genotoxic activity with unparalleled easiness, cross laboratory robustness and positive predictivity for mutagenic carcinogens (Gatehouse et al. 1994; Kirkland et al. 2006; Tennant et al. 1987). It appears logical that, provided the appropriate metabolic pathways are incorporated, the ability of a chemical to produce DNA damage that results in mutations as an essential prerequisite for tumour induction is most easily measured in bacteria.

2. The in vitro micronucleus test is endorsed as an alternative option to the in vitro chromosomal aberration test and the mouse lymphoma tk assay.

Many years of protocol evaluation and validation exercises imply that the in vitro micronucleus test has reached a status of reliability that is comparable with the mouse lymphoma tk assay in L5179Y cells or the chromosomal aberration test with various cell lines or primary human lymphocytes (Corvi et al. 2008; Lorge et al. 2007; OECD 2010). Hence, it can be used interchangeably with these assays in the regulatory world. Since many industrial laboratories already screen for genotoxic activity in the in vitro micronucleus test in early stages of non-regulatory activity, this should now enable a seamless transition from early non-GLP screening activities into the regulatory GLP testing phase.

3. An extensive review of exposure data to pharmaceuticals indicates that testing to a concentration of 1 mM (instead of 10 mM) for non-toxic compounds in mammalian cells in vitro is sufficient.

Traditionally, in vitro tests for genotoxicity have been viewed as hazard identification tests to be followed up by in vivo tests for risk characterisation or risk assessment. Under such a view, a maximal sensitivity approach has often been the goal for in vitro tests, and an upper limit of 10 mM (or 5 mg/ml) for test material in the cell culture has been applied for testing of compounds that were non-toxic. It is understood that this level represented worst-case assumptions of exposure to foreign material in vivo, for example high local exposure in the GI tract, or maximal bioavailability and thus high systemic exposure. This upper limit was also derived from early testing experience that some mutagenic carcinogens appeared to require such high concentrations to elicit a chromosome-damaging response in mammalian cells in vitro (Scott et al. 1991). Consequently, very often, positive results in vitro were viewed as potentially relevant for a chronic low-level exposure in vivo because of the stochastic element in genetic toxicology and mutation induction where fully linear dose–response

characteristics are thought to prevail. Over the years, however, there was growing evidence in the applied science for non-linearity of many aspects of genotoxic activity. Since we are interested in the risk of human exposure to relatively low doses for rather long periods, high doses in vitro had been justified on the basis that one should try to compensate for the short-term exposure in genotoxicity tests by maximising the dose. However, this often means reaching conditions of concentration and toxicity that would never occur in vivo and triggering indirect mechanism of genotoxicity (Pottenger et al. 2007). This is particularly true for pharmaceuticals, the exposure to which rarely reaches such high levels. Thus, mistrust has been building up in the judgement of in vitro positive findings. In practice, evidence for genotoxic activity in vitro has led in numerous cases to the conduct of extensive in vivo evaluations without ever reaching a conclusion that was acceptable to regulatory review. Another consequence was the discontinuation of development of potentially useful products very early before wasting resources on further activities with uneconomical delays. An essential element of in vivo testing and risk assessment is the comparison of concentrations that are positive in vitro with the exposure that can be reached under in vivo conditions. In this context, human pharmaceuticals offer the best possible judgement basis as exposure in animals and in humans is measured by default and into high dose ranges usually employed in animal studies.

A limit of 1 mM maintains the element of hazard identification, being higher than clinical exposures to known pharmaceuticals (Hardman et al. 2001), including those that concentrate in tissues and is also higher than the levels generally achievable in preclinical studies in vivo. Certain drugs are known to require quite high clinical exposures, for example, nucleoside analogues and some antibiotics (Hardman et al. 2001). While comparison of potency with existing drugs may be of interest to sponsors, perhaps even above the 1 mM limit, it is ultimately the in vivo tests that determine relevance for human safety.

4. Concerns over growing numbers of non-relevant positive findings in mammalian cell tests in vitro will also be counterbalanced by limiting the levels of cytotoxicity to ‘at most 50%’ for in vitro chromosomal aberration and micronucleus tests, that is, it is not necessary to test at more than 50% cytotoxicity. This proposal is supported by an extensive review of results obtained with in vitro hazard identification testing and in vivo risk assessment testing in the pharmaceutical industry.

Although some genotoxic carcinogens are not detectable with in vitro genotoxicity assays unless the concentrations tested induce some degree of cytotoxicity, particularly when measured by colony-forming assays, DNA-damaging agents are generally detectable with only moderate levels of toxicity (e.g. 30% reduction in growth measured at the time of sampling in the chromosomal aberration assay) (Greenwood et al. 2004). As cytotoxicity increases, mechanisms other than direct DNA damage by a compound or its metabolites can lead to ‘positive’ results that are related to cytotoxicity and not genotoxicity. Such indirect induction of DNA damage, secondary to damage to non-DNA targets, is more likely to occur above a certain concentration threshold. The disruption of cellular processes is not expected to occur at lower, pharmacologically relevant concentrations.

In cytogenetic assays, even weak clastogens that are known to be carcinogens are positive without exceeding a 50% reduction in cell counts. On the other hand, compounds that are not DNA damaging, mutagenic or carcinogenic can induce chromosome breakage at toxic concentrations. For cytogenetic assays in cell lines, measurement of cell population growth over time (by measuring the change in cell number during culture relative to control, e.g. by the method referred to as population doubling) has been shown to be a useful measure of cytotoxicity, as it is known that cell numbers can underestimate toxicity (Kirkland et al. 2007a). For lymphocyte cultures, an inhibition of mitotic index (MI) not exceeding about 50% is considered sufficient.

For the in vitro micronucleus assay, a limit of about 50% cytotoxicity is also appropriate. Moreover, for the in vitro micronucleus assay, since micronuclei are scored in the interphase subsequent to a mitotic division, it is important to verify that cells have progressed through the cell cycle. This can be done by use of cytochalasin B to allow nuclear division but not cell division, so that micronuclei can be scored in binucleate cells (the preferred method for lymphocytes). Other methods to demonstrate cell proliferation, including cell population growth over time (PD), as described above, may be used for cell lines (Kirsch-Volders et al. 2003a, b; Lorge et al. 2008; Fellows et al. 2008).

For the mouse lymphoma tk+/− assay (MLA), appropriate sensitivity is achieved by limiting the top concentration to one with close to 20% relative total growth (RTG) both for soft agar and for microwell methods. This is based on reviews of published data using the current criteria described by Moore et al. (2006), which found very few chemicals that were positive in MLA only at concentrations with less than 20% RTG and that were rodent carcinogens, and convincing evidence of genotoxic carcinogenesis for this category is lacking. The consensus is that caution is needed in interpreting results when increases in mutation are seen only below 20% RTG, and a result would not be considered positive if the increase in mutant fraction occurred only at ≤10% RTG (Moore et al. 2006, 2007). Because of the inherent difficulties to obtain an almost exact value of 20% RTG in an MLA, it is acceptable to approach a range of 10–20% to RTG for a valid assay with a compound that produces toxicity.

In conclusion, caution is appropriate in interpreting positive results obtained as reduction in growth/survival approaches or exceeds 50% for cytogenetic assays or 80% for the MLA. It is acknowledged that the evaluation of cells treated at these levels of cytotoxicity/clonal survival may result in greater sensitivity, but bears an increased risk of non-relevant positive results (Kirkland et al. 2007a). The battery approach for genotoxicity is designed to ensure appropriate sensitivity without the need to rely on single in vitro mammalian cell tests at high cytotoxicity. To obtain an appropriate toxicity range, a preliminary range-finding assay over a broad range of concentrations is useful, but in the genotoxicity assay, it is often critical to use multiple concentrations that are spaced quite closely (less than square root of 10 dilutions). Extra concentrations may be tested but not all concentrations will need to be evaluated for genotoxicity. It is not intended that multiple experiments be

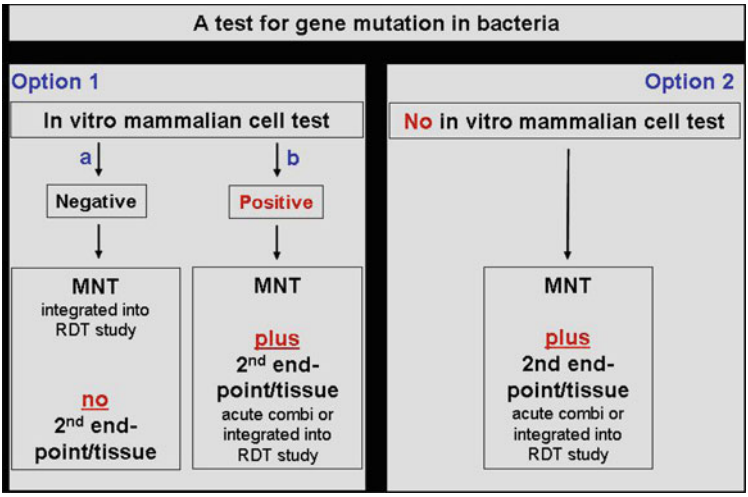


Fig. 6.2 The two options for a standard test battery of pharmaceuticals for genotoxicity according to ICH S2(R1)

carried out to reach exactly 50% reduction in growth, for example, or exactly 80% reduction in RTG.

5. As pharmaceuticals are normally tested for toxicity in rodent repeat-dose toxicity tests and as there is no longer a requirement for an acute high-dose rodent toxicity test, the assessment of genotoxicity (e.g. bone marrow micronucleus test or other tissue/endpoint) should be integrated, if feasible, into the rodent repeat-dose toxicity study to optimise animal usage.
6. The options for a standard battery of genotoxicity tests are expanded by the possibility to choose to conduct an in vivo test with investigation of genotoxic damage in two tissues instead of conducting an in vitro test with mammalian cells followed by an in vivo test (see Fig. 6.2).

In conjunction with the respective ICH carcinogenicity guidances (see under http://www.ich.org), the ICH genotoxicity guidances are setting a new standard for genetic toxicology and carcinogenesis testing, assessment and interpretation which is applicable in most parts of the world.

6.5.3 Status of ICH S2(R1): Lessons Learned About ICH and Other Processes

Comments on the above specified principles of the ICH S2(R1) Step 2 guideline were obtained by the group. These centred on the issue whether a reduction in the top concentration for in vitro tests is justified and whether an option for a test battery

without mammalian cell tests in vitro but extended in vivo testing would be endorsed. Key opposing positions were taken by US regulatory scientists involved in advising the FDA on how to test and assess pharmaceuticals for genotoxicity. In early 2009, some regulatory experts from several areas of the US FDA, EPA and USDA (US Department of Agriculture) published a review on 'Current and Future Application of Genetic Toxicity Assays: The Role and Value of In vitro Mammalian Assays' (Elespuru et al. 2009). Essentially, this group took the position that the changes introduced by the new ICH S2(R1) guideline were not thoroughly and independently reviewed, would violate good and sound principles of testing that served well over decades and replace them with not yet validated concepts and assays unsuited for regulatory review. They conclude: 'Thus, it is appropriate to apply a prudent approach to risk assessment, maintaining current testing standards that are working properly until others have proven superior by rigorous scientific evidence and widespread agreement' (Elespuru et al. 2009). Thus, in the ICH expert working group meeting in mid-2009, the situation culminated in an impasse as the FDA representative was told to hold back his signature to Step 4 of the new ICH S2(R1) guidance. This situation occurred despite widespread agreement in other regulatory agencies involved in ICH and the industry with the ICH S2(R1) process and outcome. As such, this type of situation is not unique to the S2 ICH topic but has occurred quite often in the ICH process in other guidelines. It shall be not forgotten that such a dissent was a major reason to split the intended single S2 guideline into two in the mid-1990s. However, because of the advanced process of a ready-to-sign Step 4 ICH S2(R1) guideline, the otherwise widespread agreement and the dissent of some members within one party, the FDA, the situation emerged on the ICH level as an issue which could only be controlled internally by the FDA in an 'issue resolution process'. To potentially resolve this internal conflict, an expert meeting was organised by the FDA in January 2010 to review the issues before a panel of experts from academia and other government organisations and regulatory authorities.

Although the majority of experts participated at the meeting agreed with the proposals by ICH EWG, the following objections were raised by the US regulatory group (scientists from FDA, EPA and USDA) against the ICH S2(R1) guidance (1) disagreement with an option for a test battery without an in vitro mammalian cell assay because of loss of sensitivity; (2) concern about discounting in vitro positive results in mammalian cells as non-relevant, or limiting the assay sensitivity by reducing the top concentration and controlling toxicity; (3) disagreement with the recommendation for a DNA strand-break assay (often the Comet assay in liver) as the second in vivo assay in a battery.

A philosophical difference between the proposals of the EWG in the ICH S2(R1) guidance and the US regulatory group relates to the value of the in vitro mammalian cell assays. The US regulatory group considered the data from the assays informative; as described by Elespuru et al. (2009), the assays should not be assessed only for their correlation with rodent carcinogenicity but should be regarded as sentinels of genotoxic effects that may indicate genetic risk and therefore merit further investigation. Therefore, the validity of a test battery without the in vitro assay (Option 1 in ICH S2(R1)) was seriously questioned. However, since Option 2 of the battery,

with two in vivo tests, is the same strategy used most commonly to follow up on a positive in vitro mammalian cell assay, 'further investigation' is in fact already done under this test option. Some of the US regulatory group lacked confidence in the reliance on in vivo tests to outweigh in vitro positive data, but in practice this is the strategy in the current ICH S2A and S2B guidances, in the FDA's own internal guidance (FDA 2006) and in strategies devised by ILSI–IVGT, IWGT and the WHO IPCS groups (Thybaud et al. 2007b; Eastmond et al. 2009; Lynch et al. 2011).

The US regulatory group strongly supports the use of the MLA in a test battery, primarily based on the wide range of genetic lesions, including recombination, that the assay can detect, relevant to events that occur in the development of tumours and in germ cell mutation. However, no well-founded examples have been brought forward of compounds that are known to be rodent carcinogens or germ cell mutagens, or human carcinogens, that are uniquely positive in the MLA. The known germ cell mutagens are positive in the Ames test and/or in vivo bone marrow chromosomal damage assays (Adler and Ashby 1989). This debate about inclusion of the MLA occurred also at the time the original ICH S2 battery was developed, and MLA was adopted as an option (but not a required test) in the battery because it detected a similar range of compounds to the in vitro chromosomal damage assay. Thus, the test for 'gene mutations' in the battery was the Ames test, and this has not changed in the revised ICH S2(R1) battery.

There was concern from the US regulatory group about losing information from these sensitive tests although in practice the information is not used in risk assessment for human pharmaceuticals, especially when carcinogenicity data are available. The sensitivity of these tests is recognised by many genetic toxicologists as inappropriately high as they react not only to direct DNA damage, which may be relevant at low concentrations, but to factors that occur only at high concentrations or high levels of cytotoxicity. The added information from the second in vivo test in the Option 2 test battery of ICH S2(R1) is seen as very valuable, given the facts that some known carcinogens are 'missed' in the in vitro battery due to inadequate metabolism (e.g. urethane) and that appropriate tissues can be used based on knowledge of the individual drug. The strategy also aligns with the recent increasing emphasis on assessing metabolites in toxicology.

6.5.3.1 Concentration Limit for In Vitro Mammalian Cell Assays

The ICH S2(R1) guidance maintains the option of using the in vitro mammalian cell assays but attempts to reduce the number of positive results that are not relevant to in vivo conditions by reducing the 'limit dose' top concentration from 10 mM to 1 mM and by clarifying the upper limit of toxicity. This is part of a general move forward in the broader world of genotoxicity testing of all types of chemical products and is of critical importance, especially for products for which in vivo testing is not usually done early in development or is even prohibited.

For pharmaceuticals, the recommendation to reduce the limit concentration to 1 mM was based on two factors. First, industry experience and published data

showed that positive results in the range above 1 mM were typically not confirmed as indicators of genotoxic risk by assessment of ability to react with DNA and by in vivo testing. Second, the ICH EWG wished to concentrate identification of hazards on realistic indicators of potential risk under conditions of real-life exposure, so the 1 mM limit was accepted and is still considerably higher than known human exposures to pharmaceuticals and orders of magnitude above the range at which the drug interacts with its pharmacological target and with, for example, p450 enzymes involved in metabolism (see above). However, the US regulatory group objected to the 1 mM limit on the grounds that for hazard identification high concentrations were necessary. For example, Elespuru et al. (2009) cited certain rodent carcinogens that were positive in MLA only above 1 mM. A close of examination of these does not support the argument that concentrations above 1 mM are required. For example, malonaldehyde is ubiquitous in vivo as a natural product of our metabolism and thus not considered a human risk, and toluene, which Elespuru et al. (2009) refer to as a 'well-characterised carcinogen', is in the opinion of many not a carcinogen, as it is positive in only 1 of 8 carcinogenicity studies in the Berkeley database (http://potency.berkeley.edu/), and is negative in the US NTP carcinogenicity study. Similarly, further examples given in Elespuru et al. (2009) of MLA-positive compounds that would be missed with a 1 mM concentration limit are trimethylthiourea, which induces thyroid follicular cell tumours only in the female rat and is considered a non-genotoxic carcinogen acting through altered thyroid hormone balance; acrylamide, which induces micronuclei in vivo in mice (e.g. Adler et al. 1988), and DNA damage in vivo in the liver Comet assay in mice (Ghanayem et al. 2005) and rats (Rothfuss et al. 2010); and glycidamide, the epoxide metabolite of acrylamide, which is a potent Ames mutagen. The last two examples serve to emphasise the importance of considering the test battery results overall when assessing the need for certain tests or conditions; that is, compounds that are easily detectable in other tests in a routine battery need not be used to justify an individual assay or protocol, especially when the majority of genotoxins are detectable without the need for such conditions.

During the development of the ICH S2(R1) guidance, a large effort was under way (ECVAM) to assess the effective concentration in in vitro mammalian cell assays of compounds with both genotoxicity and carcinogenicity test data. The analysis has been published (Parry et al. 2010) and summarised by Kirkland and Fowler (2010). Some of the compounds reported to be positive only above 1 mM have been retested by modern protocols and found to be negative or positive at lower concentration when tested according to standards in place since the OECD 1997 guidelines (OECD 1997). The information was assessed by the IWGT working groups in 2009 in the context of suitable test concentrations for non-pharmaceuticals, and there was general agreement that a reduction in top test concentration is warranted. Since many non-pharmaceuticals, including pesticides, industrial chemicals and food additives, have lower molecular weights (e.g. in the range of 150–250), the mg/ml concentrations that correspond to 10 mM or 1 mM are lower than those for typical pharmaceuticals with MWs of

larger than 400, so it is likely that a combination of a mM and mg/ml limit will ultimately be recommended for testing non-pharmaceuticals, but there was broad acceptance of reducing the top concentration to increase the specificity of the in vitro assays (Kirkland et al. 2011; Galloway et al. 2011). For pharmaceuticals, a note was added to the ICH S2(R1) guidance to note that higher test concentrations might be needed if the molecular weight was below 200.

6.5.3.2 Toxicity Limits for In Vitro Mammalian Cell Assays

The upper limits of toxicity and the methods for measuring toxicity have long been recognised as critical factors in designing in vitro mammalian cell assays. For the MLA, the recommendations adopted in ICH S2(R1) are those of the IWGT expert reports on the assay. ICH S2(R1) EWG recommended that an upper limit of about 80% toxicity should not be exceeded. To align with the recommendations of the IWGT expert groups (Moore et al. 2002), the ICH S2(R1) document was revised to define 'about 80%' as 80–90% relative total growth (RTG). It is also stated that one should treat with caution increases in mutation in the assay that occur only when RTG is reduced by over 80% (Mitchell et al. 1997) and that increases seen only below 10% RTG (reduction of 90% or more) are not considered positive (Moore et al. 2002). For the in vitro chromosome damage assay, both chromosomal aberrations at metaphase and the in vitro micronucleus assay, a 50% toxicity limit is recommended, with appropriate methods to ensure accurate estimation of toxicity. This was based on the original intent of the IWGTP (1994) expert working group (Galloway et al. 2011) and on published and internal industry experience. More recently, large multi-laboratory collaborative trials of the in vitro micronucleus test have verified that a wide variety of micronucleus inducers are detectable in this assay without exceeding 50% reduction in growth (Kirkland 2010a).

The reports of the 2009 IWGT meetings endorse the careful assessment of cytotoxicity by methods that assess cell growth, both for the in vitro aberration assay and for the micronucleus test (Galloway et al. 2011). In the new OECD guideline for the micronucleus assay (OECD, No. 487), an upper toxicity limit of $55 \pm 5\%$ was defined as appropriate, which is somewhat more toxic than the 50% limit recommended for pharmaceuticals in the ICH S2(R1) guidance.

Internal industry experience is that the frequency of positive results in the MLA and chromosomal aberration assays, with compounds for which no mechanistic explanation is apparent, is much lower than 25–30% when toxicity is controlled to avoid exceeding 80% reduction in RTG in MLA and 50% reduction in growth as measured by population doublings (experience of a current author). This reinforces maintaining the in vitro assays as an option in the ICH S2(R1) battery and along with the recommendations now in place for data interpretation in MLA should contribute to fewer 'non-relevant' results in these assays (Moore et al. 2011) of particular importance in product testing where in vivo follow-up is not an option.

6.5.3.3 In Vivo Genotoxicity Assays

The subject of suitable in vivo assays is also a matter of debate for some of the US regulatory group, who are concerned both on mechanistic grounds (DNA strand-break assays do not measure gene mutations) and because the Comet assay was not considered suitably validated for routine use. Previously, in the ICH S2A and S2B guidelines, although a list of in vivo assays was provided, the UDS (unscheduled DNA synthesis) assay was emphasised as there was most extensive experience with UDS at that time. In pharmaceutical testing it is extremely rare to see positive results in the UDS assay, which probably reflects the fact that it is used not for classical Ames test-positive DNA-damaging compounds for which it was 'validated' but as a follow-up to compounds positive in the in vitro mammalian cell assays, so the in vivo UDS answer is likely to be the 'right' answer regarding lack of in vivo genotoxic risk.

Experience with the Comet assay in vivo has grown dramatically in the last few years. The Comet assay has been the subject of several expert group reports with published recommendations on test design and data analysis over many years. In Europe it has been routinely used as a follow-up assay to in vitro positive genotoxicity assays, and some US pharmaceutical companies and contract labs also have quite extensive experience with the assay. It is now part of the recommendations for in vivo tests in schemes published by a range of groups of genetic toxicology experts including IPCS (Eastmond et al. 2009).

Certainly with the Comet assay in vivo, like all assays, it requires experience and attention to technical detail to obtain reliable and reproducible results, as evidenced by the validation trial initiative by JaCVAM trials with ECVAM, and ICCVAM/NICEATM is now close to the final stage. During the ICH S2(R1) deliberations, a multi-laboratory trial was done in experienced laboratories largely to assess the relative sensitivities of acute and repeat-dose protocols (Rothfuss et al. 2010), and a subsequent IWGT working group assessed this and other data on the value of the assay, the appropriate dose selection and the suitability for integration into toxicity studies, or combination with the micronucleus assay. The IWGT report (Rothfuss et al. 2011) essentially endorses the advice in the ICH S2(R1) guidance.

Lessons learned, yet again, from this process are that while scientists in the pharmaceutical industry may have extensive internal experience with assays and strategies, giving them confidence in making recommendations for change, broad acceptance of proposed changes cannot be achieved without systematic evaluation of all available data by the broader community of genetic toxicologists. For example, experienced companies knew that in developing weight of evidence, compounds positive in in vitro mammalian cell assay at concentrations above 1 mM and toxicities approaching 50% or more were unlikely to be verified as a true genotoxic risk based on studies of DNA damage and in vivo genotoxicity studies. However, while good case studies are published, much of the information is proprietary, and not many compounds progress as far as in vivo testing or carcinogenicity testing. The shared experiences with well-researched examples, together with the published data, mean that the preponderance of evidence supports changing the upper limits

for the assays. While retrospective analyses of years of published data such as those done for ECVAM (Parry et al. 2010; Kirkland and Fowler 2010) and by Kirkland et al. (2005, 2006) are limited by the fact that modern test protocols and data interpretation would alter the conclusions of the published studies, it is not practical to wait to change guidance until prospective, well-controlled multi-laboratory trials are done on every aspect of a protocol and including every ICH region. For example, the recommendations of expert groups developing the OECD guideline for the in vitro micronucleus test were based on a considerable amount of experience and published data from extensive trials, largely in the EU, but Elespuru et al. (2009) considered it not sufficiently validated, and recommendations based on this information by IWGT and the expert group that prepared the draft OECD guideline on the in vitro micronucleus assay (2006) were not accepted by the US regulatory group until a new multi-laboratory trial was done in which they had direct input into compound selection and data review (Kirkland 2010a). The OECD guideline 487 was subsequently finalised in 2010, after close to 10 years of discussion.

It is notable also that principles of testing are not universal. Early in drug development, even before any exposure of humans, there is far more information on properties of a pharmaceutical, its metabolism and distribution and its effects throughout the whole animal than there is for many other types of products, and this can be taken into account in assessing risk. Eventually, the FDA agreed to endorse the ICH S2(R1) guideline in November 2011, and thus, it is put into practice now in all ICH regions. However, it had to be made clear that this guideline would only apply to human pharmaceuticals, and thus, for example, OECD guidelines are now undergoing a revision process.

6.6 Genotoxicity Aspects in Other ICH Guidelines

The requirements for preclinical testing and the timing of studies in relation to stages of clinical development are described in the ICH M3(R2) guidance. Currently the in vitro bacterial mutation assay (Ames test) and in vitro mammalian cell assay are required before any human clinical trials, and in vivo genotoxicity testing is typically not required for first trials in people but has to be done before phase II clinical trials. As companies adopt the use of tests such as the micronucleus assay that are integrated into toxicology studies, these data will be available before early human clinical trials. Any follow-up testing to in vitro mammalian cell positive results would have to be done before clinical testing. The ICH M3(R2) guideline has some alternative recommendations for genotoxicity testing to support abbreviated clinical trials that are very short or use low doses. Genotoxicity testing is not required for very low dose initial clinical trials, such as ‘microdosing’ or PET tracer studies, provided the total dose (over multiple administrations) is less than 500 μg and the dose given is also ≤1/100th NOAEL (no adverse effect level in animal studies) and ≤1/100th of the pharmacologically active dose. Abbreviated genotoxicity test batteries are also acceptable for short-term exploratory clinical studies with limited

dose and duration. For example, when the drug is given only once and the top dose is quite limited (ICH M3(R2), approach 3), only an Ames test is required; for more extended exploratory clinical trials, an Ames test and a test for chromosome damage either in vitro or in vivo are acceptable [ICH M3(R2) approaches 4 and 5]. Because of the risk–benefit decisions involved in treating cancer patients, the ICH S9 guideline on nonclinical evaluation of anticancer pharmaceuticals notes that genotoxicity studies are not required to support clinical studies in patients with advanced cancer, but should be performed to support marketing. It also notes that in vitro positive results may not require in vivo follow-up.

Genotoxicity data are taken into account in the decision whether to test a pharmaceutical for carcinogenicity. Pharmaceuticals that are only administered for short duration often do not require carcinogenicity testing unless there is a cause for concern (ICH S1A 1995b); genotoxicity may indicate such a cause for concern, although the guideline acknowledges the need to take all information into account and notes that a single positive genotoxicity test may not indicate a hazard for people. However, 'unequivocally genotoxic' compounds are presumed to be trans-species carcinogens, indicating a potential hazard for humans, and may be treated as such without the need for carcinogenicity studies (ICH S1A).

For mechanistic/mode of action investigations, if tumours are seen in a carcinogenicity study, the possibilities for additional genotoxicity testing to help determine whether genotoxicity plays a role in tumourigenicity is discussed both in the ICH S2(R1) guidance and in ICH S1B.

The ICH S2 guidances are primarily developed for 'small molecule' pharmaceuticals. The ICH S6 guideline for preclinical safety evaluation of biotechnology-derived pharmaceuticals covers largely proteins and peptides but includes oligonucleotide products in its scope. The guideline states the standard battery of genotoxicity tests designed for typical 'small molecule' pharmaceuticals is not applicable for biopharmaceuticals, but does state that if there is a 'chemical' moiety in the molecule, such as an organic linker in a conjugated protein product, genotoxicity testing in a relevant system should be used (including newly developed systems).

Genotoxicity testing is also used in assessing the potential risks of impurities and degraded material in pharmaceuticals. The ICH Q3A and Q3B guidances include a recommendation to consider Ames and in vitro chromosome damage tests as part of 'qualification' of a drug substance or drug product if an impurity is present above defined 'qualification limits'. The Q3A and Q3B guidances acknowledge that further testing, beyond assays on the pharmaceutical containing the impurity, is needed for impurities that may be 'unusually toxic', a phrase taken to include potentially genotoxic impurities. The only published guidance on testing and controlling genotoxic impurities is that of the EU EMA (EMA 2006, 2010), and the pharmaceutical industry typically follows the recommendations of the EMA guidance together with those made in a pharmaceutical industry paper (Müller et al. 2006). A new ICH guidance, M7, is under development to address the extent of analytical identification and measurement of impurities, the assessment of structures for potential mutagenicity and biological testing of the compounds, the timing of these procedures during

clinical development and acceptable levels of exposure to mutagenic impurities of different chemical classes and for different lengths of therapy.

6.7 The IWGT Process

There is no doubt that ICH guidelines on genotoxicity testing of pharmaceuticals do not exist in isolation but have to be seen in the landscape of general scientific developments, other (national) regulatory processes and guidelines for products other than pharmaceuticals. Yet, the most important non-regulatory process of advancing science in the context of regulatory genotoxicity testing, strategy and interpretation in the so-called International Workshop on Genotoxicity Testing (IWGT), IWGTs shall be mentioned here. The IWGTs had major impact on the ICH guidelines as well as facilitated acceptance of the scientific and safety testing principles of the ICH guidelines beyond the borders of testing of pharmaceuticals.

Five workshops have been organised so far under the auspices of the International Workshop on Genotoxicity Testing, with the last one having been conducted in the summer of 2009 (Kirkland 2010b; Kirkland et al. 2011). The International Association of Environmental Mutagen Societies (IAEMS) formalised these workshops and has since held them on a continuing basis in conjunction with the International Conferences on Environmental Mutagens (ICEM) every 4 years (Kirkland et al. 2007b, c). In this way, an ongoing process of international scientific discussion and harmonization of testing methods and testing approaches has been established that can take advantage of the international experts who attend these meetings. These ongoing workshops have been demonstrated to be useful to ensure that different recommendations for methodology in these new assays do not arise in different parts of the world (Kirkland et al. 2007b, c, 2011) and thus avoid situations that could lead to the following (1) unnecessary duplication of testing to satisfy local requirements, (2) variations in the test performance, (3) potential differences in test outcome and (4) unjustified differences in the use of test data for description, assessment and management of risk. These are essentially similar concepts to the ones that drive the ICH guideline process.

The IWGT process is implemented through working groups of recognised international experts from industry, academia and the regulatory sectors, with due attention to geographical, disciplinary and sector balance. For each working group, a chairperson, deputy chair and rapporteur are appointed. Experts in the science of each topic are invited to bring experimental data to bear on the discussions; the remit of each group is to derive recommendations based on data and not on unsupported opinion or anecdotal information (Kirkland et al. 2007c). There are several objectives sought in bringing together representatives from around the world to share their experiences in generating and evaluating genotoxicity data from a variety of methodological and strategic approaches. The IWGT strives to (1) attain a greater understanding of true test performance from a wide database, (2) provide recommendations that minimise misinterpretation, (3) recognise that no single assay can detect

every genotoxicant and (4) achieve compromise for the sake of harmonization or acceptance that more than one approach is both reasonable and valid.

Because of the IWGT approach, in particular development of data-driven consensus by the key global experts from academia, government and industry, IWGT recommendations have been seen as state of the art and have high credibility. These recommendations serve as important supplements to established regulatory guidelines and provide a sound basis for updating those guidelines as the state of science advances.

With OECD and ICH guidelines constituting the two major sets of internationally harmonised genotoxicity guidelines in regulatory use, the IWGT process and working group recommendations are of particular help in supplementing test design and interpretation of genotoxicity test packages that are based on these guidelines. For further information on the IWGT recommendations, the reader is referred to the various special issues of Mutation Research that have emerged from the IWGT workshops (Kirkland et al. 2000, 2007c, 2011).

6.8 Future Perspectives in Genetic Toxicology

Future will tell whether the rather new ICH S2(R1) guideline will result in less debated assessments for genotoxicity of pharmaceuticals. It will be also interesting to see to which extent genotoxicity endpoints are becoming integrated into repeat-dose toxicity testing and how many petitioners will change their strategy in favour of omission of tests with mammalian cells in vitro in their test battery. One of the main unsolved areas of debate in genotoxicity, however, is the appropriate process for risk assessment and in this process one of the major questions is whether the underlying mechanism has a threshold, and thus, a safe level of exposure can determined or not. Presence or absence of such knowledge can drive regulatory risk management into opposite directions. The concept of thresholds has been discussed above and is rather widely accepted for genotoxicity related to indirect effects where the target is, for example, the mitotic spindle (e.g. Kirsch-Volders et al. 2009). Yet also thresholds in the dose response for chemicals that can react directly with DNA have been discussed for many years (e.g. Ehling et al. 1983). In testing practice, there is now quite a large body of evidence for thresholds for certain mutagens (Kirkland and Müller 2000; Müller and Kasper 2000). Further, a recent incident related to an accidental contamination of a pharmaceutical with ethyl methanesulfonate led to extensive data generation on several in vivo genotoxicity endpoints. With detailed exposure assessment using covalent binding data, it was possible to demonstrate that there was no risk of mutation in vivo below a surprisingly high dose of ethyl methanesulfonate (Müller et al. 2009).

Conspicuous by its absence in the ICH S2A and B and ICH S2(R1) guidances is advice on follow-up testing to a positive Ames test. Genotoxicity test follow-up strategies for chemicals in general do include the concept of following up bacterial mutagens (Eastmond et al. 2009), but for pharmaceuticals, drugs that are positive in

the Ames test are not usually developed, except when the risk–benefit justifies it for severe disease, often with short life expectancy. There are only rare circumstances where extensive testing has been done, for example, a case where it was demonstrated that a bacterial-specific enzymatic processing is responsible for mutagenesis (Suter et al. 2002); testing even included mutagenicity in the GI tract to ensure that gut bacteria would not contribute to mutagenicity in vivo. The reason for this weight on the Ames test is the high positive predictivity of the Ames test for rodent carcinogenicity. There has not been a great debate on 'how much information does it take to outweigh a positive Ames test', but in practice, the answer has been 'a negative carcinogenicity study'. Important information that would be needed to assess the significance of a positive Ames test would include knowledge of the genotoxicity and carcinogenicity of the chemical class, but this is often lacking for pharmaceuticals that are unique or part of a very small class of molecules. In vivo testing for mutation would be an important step in generating weight of evidence to follow up a positive Ames test, and equally critical is generating detailed information on the shape of the dose response for genotoxicity in vivo, since there is an increasing body of evidence that even DNA-damaging compounds may have a threshold below which genotoxicity does not occur at meaningful levels (e.g. Gocke and Müller 2009; Müller et al. 2009; Lynch et al. 2011).

The currently available assays for induction of mutations in vivo in endogenous genes such as the *hprt* gene (e.g. Skopek et al. 1995) or transgenes such as in the 'lac operon', CII or *gpt* delta (Thybaud et al. 2003) are quite laborious assays, and recommended protocols for the transgenic mouse assays include treating for at least 28 days. An in vivo assay for mutation that is more easily applicable is the 'Pig-a' assay (reviewed by Peruzzi et al. 2010), in which mutation is seen as a loss of a cell surface 'anchor' (glycosylphosphatidylinositol) and has been most widely applied to circulating red blood cells. As the assay is amenable to flow cytometry, it looks like a promising assay, being a relatively easy in vivo test to perform. There is a substantial set of information available on extensive trials to demonstrate the types of mutagens to which it is responsive, the appropriate protocols for length of treatment (e.g. Phonethepswath et al. 2010), and more importantly work in nucleated cell types has confirmed that the cell surface change is associated with DNA sequence changes in the gene itself, for example, in bone marrow cells (Kimoto et al. 2011). Currently, efforts are under way to increase the sample size in the assay to ensure it has the statistical power to detect rare events with appropriate sensitivity. If the 'Pig-a' assay fulfils its promise, it will be valuable in several applications in genetic toxicology testing, including assessing the shape of the dose response and advancing our knowledge of the range of chemical mutagens that may have thresholds and for providing weight of evidence on the in vivo effects of bacterial mutagens. Negative results in an in vivo 'Pig-a' mutation assay alone might not be enough to establish the safety of a pharmaceutical that is mutagenic in the Ames test, but might be an appropriate follow-up to a low-level impurity that is mutagenic, for example, to help establish safe levels of exposure to the impurity.

It is also difficult to determine the risk associated with positive in vivo genotoxicity test results. Some sources of false-positive results are known, such as the

increases in micronuclei associated with altered haematopoiesis (summarised in Tweats et al. 2007b). The key is related to understanding mechanism, understanding the shape of the dose response and whether there is a threshold or not. International discussions of the amount and type of evidence needed to establish such a threshold are ongoing (e.g. Thybaud et al. 2011), and it is likely that data from the 'Pig-a' mutation assay would play a role in answering such questions.

Acknowledgements This chapter is partly based on previous review articles on ICH guidelines for genotoxicity (Müller et al. 1999; MacGregor et al. 2000; Müller and Martus 2010) to ensure consistency.

References

Adler ID, Ashby J (1989) The present lack of evidence for unique germ cell mutagens. Mutat Res 212:55–66

Adler ID, Ingwersen I, Kleisch U, El Tarras A (1988) Clastogenic effects of acrylamide in mouse bone marrow cells. Mutat Res 206:379–385

Albanese R, Mirkova E, Gatehouse D, Ashby J (1988) Species-specific response to rodent carcinogens, 1,2-dimethylhydrazine and 1,2-dibromochloropropane in rodent bone marrow micronucleus assays. Mutagenesis 3:35–38

Ames B, McCann J, Yamasaki E (1975) Methods for detecting carcinogens and mutagens with the Salmonella/mammalian-microsome mutagenicity test. Mutat Res 31:347–364

Ashby J, Waters MD, Preston J, Adler I-D, Douglas GR, Fielder R, Shelby MD, Anderson D, Sofuni T, Gopalan HN, Becking G, Sonich-Mullin C (1996) IPCS harmonization of methods for the prediction and quantification of human carcinogenic/mutagenic hazard, and for indicating the probable mechanism of action of carcinogens. Mutat Res 352:153–157

Auerbach C, Robson JM (1946) Chemical production of mutations. Nature 157:302

Auletta AE, Dearfield KL, Cimino MC (1993) Mutagenicity test schemes and guidelines: U.S. EPA, Office of Pollution Prevention and Toxics and Office of Pesticide Programs. Environ Mol Mutagen 21:38–45

Bergmann K, Müller L, Teigen S (1996) The genotoxicity and carcinogenicity of paracetamol - a regulatory (re)view. Mutat Res 349:263–288

Broschinski L, Madle S, Hensel C (1998) Genotoxicity tests for new chemicals in Germany: routine in vitro systems. Mutat Res 418:121–129

Brusick D (1987) Screening chemicals for genotoxic properties. In: Principles of genetic toxicology, 2nd edn. Plenum, New York, pp 79–117

Cimino MC (2006) Comparative overview of current international strategies and guidelines for genetic toxicology testing for regulatory purposes. Environ Mol Mutagen 47:362–390

Corvi R, Albertini S, Hartung T, Hoffmann S, Maurici D, Pfuhler S, van Benthem J, Vanparys P (2008) ECVAM retrospective validation of in vitro micronucleus test (MNT). Mutagenesis 23(4):271–283

CPDB (2007) Carcinogenic Potency Data Base. http://potency.berkeley.edu/cpdb.html

CSGMT - The Collaborative Study Group for the Micronucleus Test, CSGMT (1986) Sex difference in the micronucleus test. Mutat Res 172:151–183

Dearfield KL, Moore MM (2005) Use of genetic toxicology information for risk assessment. Environ Mol Mutagen 46(4):236–245

DH (1989) Report on health and social subjects, 35. Guidelines for the testing of chemicals for mutagenicity. Committee on Mutagenicity of Chemicals in Food, Consumer Products and the Environment. Her Majesty's Stationery Office, London

DHEW (1977) U.S. Department of Health Education and Welfare (DHEW) Subcommittee on Environmental Mutagen, approaches to determining the mutagenic properties of chemicals: risk to future generations. Committee to Coordinate Toxicology and Related Programs, U.S. Department of Health Education and Welfare (DHEW), pp 1–58

DHSS (1981) Guidelines for the testing of chemicals for mutagenicity. Report on health and social subjects No. 24. Prepared by the Committee on Mutagenicity of Chemicals in Food, Consumer Products and the Environment. Department of Health and Social Security. Her Majesty's Stationery Office, London

DOH (1991) Genotoxicity assays recommended in the Health Protection Branch Genotoxicity Committee Report. Health Protection Branch, Canada Department of National Health and Welfare, Ottawa, Canada

Eastmond DA, Hartwig A, Anderson D, Anwa WA, Cimino MC, Dobrev I, Douglas GR, Nohmi T, Phillips DH, Vickers C (2009) Mutagenicity testing for chemical risk assessment: update of the WHO/IPCS harmonized scheme. Mutagenesis 24:341–349

EC (2008) Annex V. Part B: Methods for the determination of toxicity. Directive 67/548/EEC, http://ecb.jrc.ec.europa.eu/testing-methods/annex5/

EEC (1987) Notes for guidance for the testing of medicinal products for their mutagenic potential. Official Journal European Community, L73

Ehling UH, Macheme L, Buselmaier W, Dýck J, Frohberg H, Kratochvilova J, Lan R, Lorke D, Müller D, Peh J, Röhrborn G, Rol R, Schulze-Schencking M, Wiemann H (1978) Standard protocol for the dominant lethal test on male mice set up by the work group "Dominant Lethal Mutations of the ad hoc Committee Chemogenetics". Arch Toxicol 39:173–185

Ehling UH, Averbec D, Cerutti PA, Friedman J, Greim HJ, Kolbye AC Jr, Mendelsohn ML (1983) ICPEMC Publication no. 10. Review of the evidence for the presence or absence of thresholds in the induction of genetic effects by genotoxic chemicals. Mutat Res 123:281–341

Elespuru RK, Agarwal R, Atrakchi AH, Bigger CAH, Heflich RH, Jagannath D, Levy D, Moore MM, Ouyang Y, Robison TW, Sotomayo RE, Cimino MC, Dearfield KL (2009) Current and future application of genetic toxicity assays: the role and value of in vitro mammalian assays. Toxicol Sci 109:172–179

EMA (2006) Guideline on the limits of genotoxic impurities. Committee on Human Medicinal Products; Safety Working Party (SWP) CPMP/SWP/5199/02. http://www.ema.europa.eu/ema/index

EMA (2010) Questions and answers on the "Guideline on the limits of genotoxic impurities". Committee on Human Medicinal Products; Safety Working Party (SWP) CHMP/SWP/431994/2007 Rev. 3. http://www.ema.europa.eu/ema/index

EPA (2008) OPPTS harmonized test guidelines Series 870 Health effects test guidelines. http://www.epa.gov/opptsfrs/publications/OPPTS_Harmonized/870_Health_Effects_Test_Guidelines/Series/

FDA (2006) Recommended approaches to integration of genetic toxicology study results. http://www.fda.gov/cder/guidance/index.htm

Fellows MD, O'Donovan MR, Lorge E, Kirkland D (2008) Comparison of different methods for an accurate assessment of cytotoxicity in the in vitro micronucleus test: II: Practical aspects with toxic agents. Mutat Res 655:4–21

Galloway S, Lorge E, Aardema MJ, Eastmond D, Fellows M, Heflich R, Kirkland D, Levy DD, Lynch AM, Marzin D, Morita T, Schuler M, Speit G (2011) Workshop summary: Top concentration for in vitro mammalian cell genotoxicity assays; and report from working group on toxicity measures and top concentration for in vitro cytogenetics assays (chromosome aberrations and micronucleus). Mutat Res 723:77–83

Gatehouse D, Haworth S, Cebula T, Gocke E, Kier L, Matsushima T, Melcion C, Nohmi T, Ohta T, Venitt S, Zeiger E (1994) Report from the working group on bacterial mutation assays: international workshop on standardisation of genotoxicity test procedures. Mutat Res 312:217–233

Ghanayem BI, Witt KL, Kissling GE, Tice RR, Recio L (2005) Absence of acrylamide-induced genotoxicity in CYP2E1-null mice: evidence consistent with a glycidamide-mediated effect. Mutat Res 578:284–297

Gocke E, Müller L (2009) In vivo studies in the mouse to define a threshold for the genotoxicity of EMS and ENU. Mutat Res 678:101–107

Greenwood SK, Hill RB, Sun JT, Armstrong MJ, Johnson TE, Gara JP, Galloway SM (2004) Population doubling: a simple and more accurate estimation of cell growth suppression in the *in vitro* assay for chromosomal aberrations that reduces irrelevant positive results. Environ Mol Mutagen 43:36–44

Hardman JG, Limbird LE, Gilman AG (eds) (2001) Goodman and Gilman's the pharmacological basis of therapeutics, 10th edn. McGraw-Hill Professional, New York

Holden HE (1982) Comparison of somatic and germ cell models for cytogenetic screening. J Appl Toxicol 2:196–200

Hollaender A (1971) Chemical mutagens: principles and Methods for their detection. Plenum, New York

Honma M, Hayashi M, Shimada N, Tanaka S, Wakuri T, Awogi KI, Yanamoto N-U, Kodani Y, Nishi M, Nakadate T, Sofuni T (1999) Evaluation of the mouse tk assay (microwell method) as an alternative to the in vitro chromosomal aberration test. Mutagenesis 24:2–22

ICH (1995b) ICH S1A: Need for carcinogenicity testing of pharmaceuticals. http://www.ich.org

ICH (1995a) ICH S2A: Guidance on specific aspects of regulatory genotoxicity tests for pharmaceuticals. http://www.ich.org

ICH (1997) ICH S2B: A standard battery for genotoxicity testing of pharmaceuticals. http://www.ich.org

ICH (2008) ICH S2(R1): Guidance on genotoxicity testing and data interpretation for pharmaceuticals intended for human use. http://www.ich.org

Jacobs A (2005) Prediction of 2-year carcinogenicity study results for pharmaceutical products: how are we doing? Toxicol Sci 88:18–23

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

JMHW (1984) Information on guidelines of toxicity studies required for applications for approval to manufacture (import) drugs (Part 1). Notification No 118 of the Pharmaceutical Affairs Bureau. Ministry of Health and Welfare, Tokyo, Japan

Kasper P, Uno Y, Mauthe R, Asano N, Douglas G, Matthews E, Moore M, Müller L, Nakajima M, Singer T, Speit G (2007) Follow-up testing of rodent carcinogens not positive in the standard genotoxicity testing battery: IWGT workgroup report. Mutat Res 627:106–116

Kimoto T, Suzuki K, Kobayashi XM, Dobrovolsky VN, Heflich RH, Miura D, Kasahara Y (2011) Manifestation of Pig-a mutant bone marrow erythroids and peripheral blood erythrocytes in mice treated with N-ethyl-N-nitrosourea: direct sequencing of Pig-a cDNA from bone marrow cells negative for GPI-anchored protein expression. Mutat Res 723:36–42

Kirkland DJ (1993) Genetic toxicology testing requirements: official and unofficial views from Europe. Environ Mol Mutagen 21:8–14

Kirkland D (2010a) Evaluation of different cytotoxic and cytostatic measures for the in vitro micronucleus test (MNVit): introduction to the collaborative trial. Mutat Res 702:135–138

Kirkland D (2010b) Evaluation of different cytotoxic and cytostatic measures for the in vitro micronucleus test (MNVit): summary of results in the collaborative trial. Mutat Res 702:139–147

Kirkland D, Fowler P (2010) Further analysis of Ames-negative rodent carcinogens that are only genotoxic in mammalian cells in vitro at concentrations exceeding 1 mM, including retesting of compounds of concern. Mutagenesis 25:539–553

Kirkland D, Müller L (2000) Interpretation of the biological relevance of genotoxicity test results: the importance of thresholds. Mutat Res 464:137–147

Kirkland D, Hayashi M, MacGregor JT, Müller L, Schechtman L, Sofuni T (2000) Summary of major conclusions from the international workshop on genotoxicity test procedures (IWGTP). Environ Mol Mutagen 35:162–166

Kirkland D, Aardema M, Henderson L, Müller L (2005) Evaluation of the ability of three in vitro genotoxicity tests to discriminate rodent carcinogens and non-carcinogens. I. Sensitivity, specificity and relative predictivity. Mutat Res 584:1–257

Kirkland D, Aardema M, Müller L, Hayashi M (2006) Evaluation of the ability of a battery of 3 in vitro genotoxicity tests to discriminate rodent carcinogens and non-carcinogens. II. Further analysis of mammalian cell results, relative predictivity and tumour profiles. Mutat Res 608:29–42

Kirkland D, Pfuhler S, Tweats D, Aardema M, Corvi R, Darroudi F, Elhajouji A, Glatt H, Hastwell P, Hayashi M, Kasper P, Kirchner S, Lynch A, Marzin D, Maurici D, Meunier JR, Müller L, Nohynek G, Parry J, Parry E, Thybaud V, Tice R, van Benthem J, Vanparys P, White P (2007a) How to reduce false positive results when undertaking in vitro genotoxicity testing and thus avoid unnecessary follow-up animal tests. Report of an ECVAM workshop. Mutat Res 628:31–55

Kirkland D, Hayashi M, Jacobson-Kram D, Kasper P, MacGregor JT, Müller L, Uno Y (2007b) Summary of major conclusions from the 4th IWGT, San Francisco, 9–10 September, 2005. Mutat Res 627:5–9

Kirkland DJ, Hayashi M, Jacobson-Kram D, Kasper P, MacGregor JT, Müller L, Uno Y (2007c) The international workshops on genotoxicity testing (IWGT): history and achievements. Mutat Res 627:1–4

Kirkland D, Kasper P, Müller L, Corvi R, Speit G (2008) Recommended lists of genotoxic and non-genotoxic chemicals for assessment of the performance of new or improved genotoxicity tests: a follow-up to an ECVAM workshop. Mutat Res 653:99–108

Kirkland D, Hayashi M, Jacobson-Kram D, Kasper P, Gollapudi B, Müller L, Uno Y (2011) Summary of major conclusions from the 5th IWGT, Basel, Switzerland, 17–19 August 2009. Mutat Res 723:73–76

Kirsch-Volders M, Sofuni T, Aardema M, Albertini S, Eastmond D, Fenech M, Ishidate M, Kirchner S, Lorge E, Morita T, Norppa H, Surralles J, Vanhauwaert A, Wakata A (2003a) Report from the in vitro micronucleus assay working group. Mutat Res 540:153–163

Kirsch-Volders M, Sofuni T, Aardema M, Albertini S, Eastmond D, Fenech M, Ishidate M, Kirchner S, Lorge E, Morita T, Norppa H, Surralles J, Vanhauwaert A, Wakata A (2003b) Report from the in vitro micronucleus assay working group. Corrigendum. Mutat Res 564:97–100

Kirsch-Volders M, Gonzalez L, Carmichael P, Kirkland D (2009) Risk assessment of genotoxic mutagens with thresholds: a brief introduction. Mutat Res 678:72–75

Lorge E, Lambert C, Gervais V, Becourt-Lhote N, Delongeas J-L, Claude N (2007) Genetic toxicity assessment: employing the best science for human safety evaluation. Part II. Performance of the in vitro micronucleus test compared to the mouse lymphoma assay and the in vitro chromosome aberration assay. Toxicol Sci 96:214–217

Lorge E, Hayashi M, Albertini S, Kirkland D (2008) Comparison of different methods for an accurate assessment of cytotoxicity in the in vitro micronucleus test. I: Theoretical aspects. Mutat Res 655:1–3

Lynch AM, Sasaki JC, Elespuru R, Jacobson-Kram D, Thybaud V, De Boeck M, Aardema MJ, Aubrecht J, Benz RD, Dertinger SD, Douglas GR, White PA, Escobar PA, Fornace A Jr, Honma M, Naven RT, Rusling JF, Schiestl RH, Walmsley RM, Yamamura E, van Benthem J, Kim JH (2011) New and emerging technologies for genetic toxicity testing. Environ Mol Mutagen 52(3):205–223

MacGregor JT (1994) Environmental mutagen: past and future directions. Mutat Res 23:73–77

MacGregor JT, Casciano D, Müller L (2000) Strategies and testing methods for identifying mutagenic risks. Mutat Res 455:3–21

Madle S, Korte A, Bass R (1987) Experience with mutagenicity testing of new drugs: viewpoint of a regulatory agency. Mutat Res 182:187–192

Maron DM, Ames BN (1983) Revised methods for the *Salmonella* mutagenicity test. Mutat Res 113:173–215

Matthews EJ, Kruhlak NL, Cimino MC, Benz RD, Contrera JF (2006a) An analysis of genetic toxicity, reproductive and developmental toxicity, and carcinogenicity data: I. Identification of carcinogens using surrogate endpoints. Regul Toxicol Pharmacol 44:83–96

Matthews EJ, Kruhlak NL, Cimino MC, Benz RD, Contrera JF (2006b) An analysis of genetic toxicity, reproductive and developmental toxicity and carcinogenicity data, II. Identification of genotoxicants, reprotoxicants and carcinogens using in silico methods. Regul Toxicol Pharmacol 44:97–110

McCann J, Choi E, Yamasaki E, Ames BN (1975) Detection of carcinogens as mutagens in the Salmonella/microsome test: assay of 300 chemicals. Proc Natl Acad Sci USA 72:5135–5139

Meselson M (1971) Preface. In: Hollaender A (ed) Chemical mutagens: principles and methods for their detection. Plenum, New York, pp 9–12

Mitchell AD, Auletta AE, Clive D, Kirby PE, Moore MM, Myhr BD (1997) The L5178Ytk$^{+/-}$ mouse specific gene and chromosomal mutation assay. A phase III report of the U.S Environmental Protection Agency Gene-Tox Program. Mutat Res 384:177–303

Moore MM, Honma M, Clements J, Harrington-Brock K, Awogi T, Bolcsfold G, Cifone M, Collard D, Fellows M, Flanders K, Gollapudi B, Jenkinson P, Kirby P, Kirchner S, Kraycer J, McEnaney S, Muster W, Myhr B, O'Donovan M, Oliver J, Ouldelhkim M, Pant K, Preston R, Riach C, San R, Shimada H, Stankowski LF Jr (2002) Mouse lymphoma thymidine kinase gene mutation assay: follow-up international workshop on genotoxicity test procedures, New Orleans, Louisiana, April 2000. Environ Mol Mutagen 40:292–299

Moore MM, Honma M, Clements J, Bolcsfoldi G, Burlinson B, Cifone M, Clarke J, Delongchamp R, Durward R, Fellows M, Gollapudi B, Hou S, Jenkinson P, Lloyd M, Majeska J, Myhr B, O'Donovan M, Omori T, Riach C, San R, Stankowski LF Jr, Thakur AK, Van Goethem F, Wakuri S, Yoshimura I (2006) Mouse lymphoma thymidine kinase gene mutation assay: follow-up meeting of the international workshop on genotoxicity testing–Aberdeen, Scotland, 2003–assay acceptance criteria, positive controls, and data evaluation. Environ Mol Mutagen 47:1–5

Moore MM, Honma M, Clements J, Bolcsfoldi G, Burlinson B, Cifone M, Clark J, Clay P, Doppalapudi R, Fellows M, Gollapudi B, Ho S, Jenkinson P, Muster W, Pan K, Kidd D, Lorge E, Lloyd M, Myhr B, O'Donovan M, Riach C, Stankowski LF Jr, Thakur AK, Van Goethem F (2007) Mouse lymphoma thymidine kinase gene mutation assay: meeting of the international workshop on genotoxicity testing, San Francisco, 2005, recommendations for 24-h treatment. Mutat Res 627:36–40

Moore MM, Honma M, Clements J, Awogi T, Douglas GR, van Goetham F, Gollapudi B, Kimura A, Muster W, O'Donovan M, Schoeny R, Wakuri S (2011) Suitable top concentration for tests with mammalian cells: mouse lymphoma assay workgroup. Mutat Res 723:85–86

Müller L, Kasper P (2000) Human biological relevance and the use of threshold arguments in regulatory genotoxicity assessment: experience with pharmaceuticals. Mutat Res 464:19–34

Müller L, Martus HJ (2010) Genetic toxicology testing guidelines and regulations. In: Hsu G, Stedeford T (eds) Cancer risk assessment: chemical carcinogenesis. From biology to standards quantification. Wiley, Hoboken, pp 238–271

Müller L, Kasper P, Madle S (1991) The quality of genotoxicity testing of drugs. Experiences of a regulatory agency with new and old compounds. Mutagenesis 6:143–149

Müller L, Kikuchi Y, Probst G, Schechtman L, Shimada H, Sofuni T, Tweats D (1999) ICH-harmonised guidances on genotoxicity testing of pharmaceuticals: evolution, reasoning and impact. Mutat Res 436:195–225

Müller L, Blakey D, Dearfield KL, Galloway S, Guzzie P, Hayashi M, Kasper P, Kirkland D, MacGregor JT, Parry JM, Schechtman L, Smith A, Tanaka N, Tweats D, Yamasaki H (2003) Strategy for genotoxicity testing and stratification of genotoxicity test results–report on initial activities of the IWGT Expert Group. Mutat Res 540:177–181

Müller L, Mauthe RJ, Riley CM, Andino MM, De Antonis D, Beels C, DeGeorge J, De Knaep AGM, Ellison D, Fagerland JA, Frank R, Fritschel B, Galloway S, Harpur E, Humfrey CDN, Jacks AS, Jagota N, Mackinnon J, Mohan G, Ness DK, O'Donovan MR, Smith MD, Vudathala G, Yotti L (2006) A rationale for determination, testing and control of genotoxic impurities in pharmaceuticals. Regul Toxicol Pharmacol 44:198–211

Müller L, Gocke E, Lave T, Pfister T (2009) Ethyl methanesulfonate toxicity in Viracept – a comprehensive human risk assessment based on threshold data for genotoxicity. Toxicol Lett 190:317–329 (special issue)

NRC (1983) Identifying and estimating the genetic impact of chemical mutagens, Committee on chemical environmental mutagens, Board on toxicology and environmental health hazards, Commission on life sciences. National Academy Press, Washington, DC, pp 1–295
OECD (1983) OECD guidelines for testing of chemicals. Genetic toxicology, No 471–474. Organisation for Economic Cooperation and Development, Paris, 26 May 1983
OECD (1984) ibid, No 475–478, 4 Apr 1984
OECD (1986) ibid, No 479–485, 23 Oct 1986
OECD (1997) Guidelines for genetic toxicology. http://www.oecd.org/dataoecd/9/11/33663321.pdf
OECD (2010) OECD guideline for the testing of chemicals. 487. In vitro micronucleus test. http://www.oecd-ilibrary.org/content/book/9789264091016-en
Parry JM, Parry E, Phrakonkham P, Corvi R (2010) Analysis of published data for top concentration considerations in mammalian cell genotoxicity testing. Mutagenesis 25:531–538
Peruzzi B, Araten DJ, Notaro R, Luzzatto L (2010) The use of PIG-A as a sentinel gene for the study of the somatic mutation rate and of mutagenic agents in vivo. Mutat Res 70:3–10
Phonethepswath S, Franklin D, Torous DK, Bryce SM, Bemis JC, Raja S, Avlasevich S, Weller P, Hyrien O, Palis J, Macgregor JT, Dertinger SD (2010) Pig-a mutation: kinetics in rat erythrocytes following exposure to five prototypical mutagens. Toxicol Sci 114:59–70
Pottenger LH, Bus JS, Gollapudi BB (2007) Genetic toxicity assessment: employing the best science for human safety evaluation. Part VI: When salt and sugar and vegetables are positive, how can genotoxicity data serve to inform risk assessment? Toxicol Sci 98:327–331
Probst G (1994) Validation of target tissue exposure for *in vivo* tests. In: D-Arcy PF, Harron DWG (eds) Proceedings of the second international conference on harmonization (ICH2). Greystoke Books, N. Ireland, pp 249–252
Rothfuss A, O'Donovan M, DeBoeck M, Brault D, Czich A, Custer I, Hamada S, Plappert-Helbig U, Hayashi M, Howe J, Kraynak A, van der Leede B, Nakajima M, Priestley C, Thybaud V, Saigo K, Sawant S, Shi J, Storer R, Struwe M, Vock E, Galloway S (2010) Collaborative study on 15 compounds in the rat liver Comet assay integrated into 2- and 4-week repeat-dose studies. Mutat Res 702:40–69
Rothfuss A, Honma M, Czich A, Aardema MJ, Burlinson B, Galloway S, Hamada S, Kirkland D, Heflich RH, Howe J, Nakajima M, O'Donovan M, Plappert-Helbig U, Priestley C, Recio L, Schuler M, Uno Y, Martus H (2011) Improvement of in vivo genotoxicity assessment: combination of acute tests and integration into standard toxicity testing. Mutat Res 723:108–120
Scott D, Galloway SM, Marshall RR, Ishidate M Jr, Brusick D, Ashby J, Myhr BC (1991) Genotoxicity under extreme culture conditions. A report from ICPEMC Task Group 9. Mutat Res 257:147–204
Shelby MD (1996) Selecting chemicals and assays for assessing mammalian germ cell mutagenicity. Mutat Res 352:159–167
Shelby MD, Sofuni T (1991) Toxicology testing requirements and the U.S.-Japan collaborative study on in vitro tests for chromosomal aberrations. Environ Health Perspect 94:255–259
Shelby MD, Zeiger E (1990) Activity of human carcinogens in the Salmonella and rodent bone-marrow cytogenetic tests. Mutat Res 234:257–261
Shirasu Y (1988) The Japanese mutagenicity studies guidelines for pesticide registration. Mutat Res 205:393–395
Skopek TR, Kort KL, Marino DR (1995) Relative sensitivity of the endogenous *hprt* gene and lacI transgene in ENU-treated Big Blue™ B6C3F1 mice. Environ Mol Mutagen 26:9–15
Snyder RD (2009) An update on the genotoxicity and carcinogenicity of marketed pharmaceuticals with reference to in silico predictivity. Environ Mol Mutagen 50:435–450
Snyder RD (2010) Possible structural and functional determinants contributing to the clastogenicity of pharmaceuticals. Environ Mol Mutagen 51:800–814
Snyder RD, Green JW (2001) A review of the genotoxicity of marketed pharmaceuticals. Mutat Res 488:151–169
Suter W, Hartmann A, Poetter F, Sagelsdorff P, Hoffmann P, Martus HJ (2002) Genotoxicity assessment of the antiepileptic drug AMP397, an Ames-positive aromatic nitro compound. Mutat Res 518:181–194

Tennant RW, Margolin BH, Shelby MD, Zeiger E, Haseman JK, Spalding J, Caspary W, Resnick M, Stasiewicz S, Anderson B, Minor R (1987) Prediction of chemical carcinogenicity in rodents from in vitro genetic toxicity assays. Science 236:933–941

Thybaud V, Dean S, Nohmi T, de Boer J, Douglas GR, Glickman BW, Gorelick NJ, Heddle JA, Heflich RH, Lambert I, Martus HJ, Mirsalis JC, Suzuki T, Yajima N (2003) In vivo transgenic mutation assays. Mutat Res 540:141–145

Thybaud V, Aardema M, Clements J, Dearfield K, Galloway S, Hayashi M, Jacobson-Kram D, Kirkland D, MacGregor JT, Marzin D, Ohyama W, Schuler M, Suzuki H, Zeiger E (2007a) Strategy for genotoxicity testing: hazard identification and risk assessment in relation to in vitro testing. Mutat Res 627:41–58

Thybaud V, Aardema M, Casciano D, Dellarco V, Embry MR, Gollapudi BB, Hayashi M, Holsapple MP, Jacobson-Kram D, Kasper P, MacGregor JT, Rees R (2007b) Relevance and follow-up of positive results in in vitro genetic toxicity assays: an ILSI-HESI initiative. Mutat Res 633:67–79

Thybaud V, MacGregor JT, Müller L, Crebelli R, Dearfield K, Douglas G, Farmer PB, Gocke E, Hayashi M, Lovell DP, Lutz WK, Marzin D, Moore M, Nohmi T, Phillips DH, Van Benthem J (2011) Strategies in case of positive in vivo results in genotoxicity testing. Mutat Res 723:121–128

Tinwell H, Ashby J (1991) Activity of the human carcinogen MeCCNU in the mouse bone marrow micronucleus assay. Environ Mol Mutagen 17:152–154

Tweats DJ, Blakey D, Heflich RH, Jacobs A, Jacobsen SD, Morita T, Nohmi R, O'Donovan MR, Sasaki YF, Sofuni T, Tice R (2007a) Report of the IWGT working group on strategy/interpretation of regulatory *in vivo* tests. II. Identification of in vivo-only positive compounds in the bone marrow micronucleus test. Mutat Res 627:92–105

Tweats DJ, Blakey D, Heflich RH, Jacobs A, Jacobsen SD, Morita T, Nohmi R, O'Donovan MR, Sasaki YF, Sofuni T, Tice R (2007b) Report of the IWGT working group on strategy/interpretation of regulatory *in vivo* tests. I. Increases in micronucleated bone marrow cells in rodents that do not indicate genotoxic hazards. Mutat Res 627:78–91

UK COM (2000) Guidance on a strategy for testing of chemicals for mutagenicity. United Kingdom Committee on Mutagenicity of Chemicals in Food, Consumer Products and the Environment, pp 1–36. http://www.dh.gov.uk/assetRoot/04/07/71/96/04077196.pdf

UK EMS (1990) Kirkland DJ (ed) Basic mutagenicity tests, UKEMS recommended procedures, UKEMS sub-committee on guidelines for mutagenicity testing. Report. Part I revised. Cambridge University Press, Cambridge, UK

UK EMS (1993) Kirkland DJ, Fox M (ed) Supplementary mutagenicity tests, UKEMS recommended procedures, UKEMS sub-committee on guidelines for mutagenicity testing. Report. Part II revised. Cambridge University Press, Cambridge, UK

US EPA (1982a) Health effects test guidelines. Office of Toxic Substances. Environmental Protection Agency, Washington, DC, Oct 1982

US EPA (1982b) Pesticide assessment guidelines. Subdivision F. Hazard evaluation: human and domestic animals. Office of Pesticides and Toxic Substances. Environmental Protection Agency, Washington, DC, Oct 1982

US FDA (1982) US Food and Drug Administration, toxicological principles for the safety assessment of direct food additives and color additives used in food. US FDA, Washington, DC

US FDA (1993) US Food and Drug Administration, Draft. Toxicological principles for the safety assessment of direct food additives and color additives used in food ("Redbook II"). US FDA, Washington, DC

US FDA (2000) Toxicological principles for the safety assessment of food ingredients. Redbook 2000. IV.C.1. Short-term tests for genetic toxicity. United States Food and Drug Administration, Center for Food Safety and Applied Nutrition, Office of Food Additive Safety, Washington, DC. http://www.cfsan.fda.gov/~redbook/red-ivc1.html

US FDA (2006) US Food and Drug Administration, Guidance for industry and review staff recommended approaches to integration of genetic toxicology study results. http://www.fda.gov/cder/guidance/6848fnl.htm

US TSCA (1976) United States Congress, Toxic Substances Control Act, Appendix. In: Hollaender A, DeSerres FJ (eds) Chemical mutagens: principles and methods for their detection, vol 5. Plenum, New York, pp 287–335

Wassom JS (1989) Origins of genetic toxicology and the environmental mutagen society. Environ Mol Mutagen 14:1–6

Waters MD, Auletta A (1981) The GEN-TOX program: genetic activity evaluation. J Chem Inf Comput Sci 21:35–38

Waters MD, Stack HF, Jackson MA, Bridges BA, Adler I-D (1994) The performance of short-term tests in identifying potential germ cell mutagens: a qualitative and quantitative analysis. Mutat Res 341:109–131

Wilcox P, Naidoo A, Wedd DJ, Gatehouse DG (1990) Comparison of *Salmonella typhimurium* TA102 with *Escherichia coli* WP2 tester strains. Mutagenesis 5:285–291

Zeiger E (2010) Historical perspective on the development of the genetic toxicology test battery in the United States. Environ Mol Mutagen 51:781–791

Chapter 7
Toxicokinetics: A Guidance for Assessing Systemic Exposure in Toxicology Studies, Where Are We Now; An S3A/S3B Update (1995–2011)

Bruce Campbell and Bob Ings

Abstract The ICH S3A and S3B guidances written in 1995 have been critically examined by one author whom helped write the original guidance and the other who has had to put them into practice. The objectives of the guidances were to clarify across the three regions when, what and how drug levels in pivotal safety studies should be measured using GLP-validated methods and how the results could be used to reduce the number of exploratory animal pharmacokinetic studies and reduce their emphasis. Although it has formed a good basic framework to move forward, subsequent guidances have been required. These have included the metabolites in safety testing (MIST) providing further information on which metabolites to measure and when disproportionate human levels not qualified in animals need to be further tested and bioanalytically validated. New advances in sampling techniques, composite and auto-sampling, microsampling with plasma separation, dried blood spot analysis (DBS) and kinetic population approaches of the data to enable serial concomitant sampling are discussed. Other topics of uncertainty which are discussed include when to measure tissue distribution (ICH S3B), should exposure measurement be included in in vitro studies (mutagenicity, hERG, cytotoxicity, etc.), should protein binding be measured at NOAEL and safety margins expressed as total or free unbound levels and when is Cmax more appropriate than AUC. It is suggested that toxicokinetic–toxicodynamic relationships should be investigated more frequently where possible using established biomarkers.

B. Campbell (✉)
Department of Pharmacology, Kings College Guys Campus London,
Proximagen Neuroscience, London, UK
e-mail: bruce.campbell@proximagen.com

B. Ings
RMI-Pharmacokinetics, Carlsbad, CA, USA

J.W. van der Laan and J.J. DeGeorge (eds.), *Global Approach in Safety Testing*, AAPS Advances in the Pharmaceutical Sciences Series 5, DOI 10.1007/978-1-4614-5950-7_7,

7.1 Introduction

It was more than 16 years ago that a group of toxicologists and pharmacokineticists were brought together in Europe to develop the ICH guidelines on toxicokinetics and how they should be used to assess the exposure of drugs in safety studies. Whereas pharmacokinetics had been a discipline used in drug development for more than 50 years, the term toxicokinetics was only coined in 1976 as a new word, when it was first translated into English from the original French language paper on Toxicocinetique (Carrera et al. 1976). Since then, the importance of measuring drug levels in safety studies to understand the extent of exposure has grown with an ever-increasing number of published papers (Fig. 7.1). It is of interest that many of the papers in the nineties around the time of the S3A guidance (ICH, 1995b) were discussing the value of toxicokinetics and how it should measured, whilst more recent papers have shown examples of how this approach can be used with practical examples. Acknowledging the simplistic nature of this survey, the number of papers published on toxicodynamics appears to be considerably less than for toxicokinetics and mostly referring to environmental toxins. Of interest, the number where both topics are discussed represents almost 50% of those on toxicodynamics alone, suggesting, perhaps, that toxicologists understand more the importance of measuring both than the kineticists.

This review attempts to look at the following: What we have learnt over the intervening years with regard to this new discipline and how useful were the original guidelines, did they provide enough detail, did they miss anything out, were topics discussed and not included and why, and what improvements could be made? Finally did the guidance help in the interpretation of safety studies and their

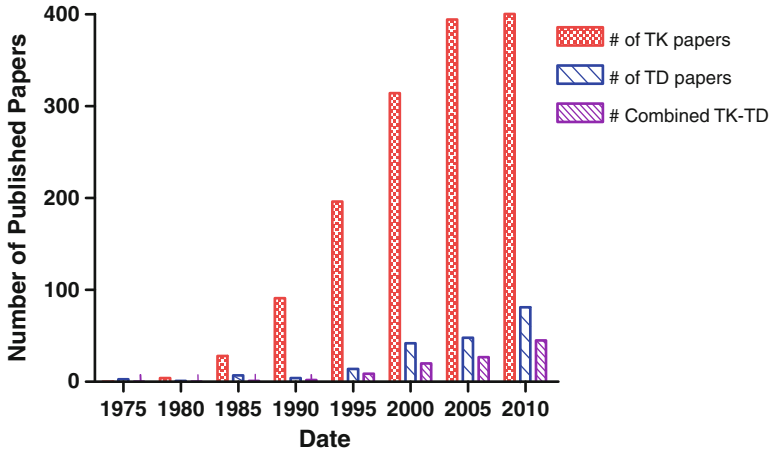


Fig. 7.1 Graph showing the number of publications mentioning toxicokinetics and toxicodynamics in the title from 1970 to 2010 (PubMed), indicating for the former that there was a large increase at the time the S3A guidance was published, but over the last 5 years, it has plateaued

relevance to clinical studies, did they save time and most importantly, one of the original objectives of the ICH process, has it reduced the number of animals used in development?

The following discussion is written with a development focus in mind, although the original template used for formulating the ICH guidelines is followed wherever possible.

7.2 General Principles and Objectives

It had long been recognised that the use of doses in terms of mg/kg comparisons between the species used in safety studies and doses used in humans was less than perfect since it was clear that the dose per kilo needed to produce effects in small animals was frequently much larger than that found to be active in humans. However, it was also observed that when plasma levels were compared between species for a given pharmacological response, in many cases, they were more similar irrespective of the doses used. Work by Boxenbaum (1982) and others has shown that the size of the animal can be an important determinant of the clearance of drugs and the resulting blood levels. With the use of allometric scaling or body surface area to extrapolate between species, better predictions of human levels can be obtained from animal data. Indeed, for many years, anticancer drugs have been dosed to humans using body surface area. Further groups (Peck et al. 1994) established the importance of understanding the integration of pharmacokinetics and toxicokinetics with all aspects of drug development and how the measurement of blood/plasma levels and their relationship with activity was paramount for the full interpretation of the action of a drug.

Although not formalised before, this understanding of the importance of plasma levels or drug exposure rather than dose per se formed the backdrop for writing the guidelines using pharmacokinetics (toxicokinetics) in the assessment and interpretation of animal safety studies and the extrapolation to humans. In particular, it was felt that although pharmaceutical companies were, in many cases, doing this already, the extent varied widely with some probably doing too much and using an unnecessarily large number of animals, whilst others were doing too little or nothing at all. Thus, the new S3 guidance (ICH, 1995b) was developed to form a framework for scientists involved in safety assessment from all three regions (the USA, Europe and Japan) to have a common document to refer to. The group involved in drawing up these guidelines was a mixture of in vivo and in vitro toxicologists, pharmacokineticists and those involved in metabolism so that all aspects, both practical and academic, could be incorporated into the guidance.

At the onset, it was agreed that this was for guidance and the use of TK must be assessed on a case-by-case basis as there may be valid reasons, albeit few, why TK may not be appropriate (e.g. non-absorbed GI tract active compounds). However, it was generally accepted that in most cases, exposure should be determined in pivotal safety studies as an integral part of the nonclinical testing programme. It should be

undertaken wherever possible and, if not, a justification given for the failure to do so. Some discussion took place on the primary objectives of toxicokinetics but its simplistic form 'to describe the systemic exposure achieved in animal studies and its relationship to dose level and time course of the toxicity study' covered all the salient points put forward by different parties. Other suggestions included, as secondary objectives, 'to try and relate the exposure to safety findings, support the choice of which species should be used for the safety studies and to the use of the findings to help design further studies'. These were all felt to be important but not necessarily essential to the primary focus of the guidance at the time of writing and, if included in any detail, would make the guidance too unwieldy, diluting the main objective. Clearly as TK has become routine, the use of the generated results in the interpretation of safety issues has become increasingly important and can make the difference between a drug getting approved or not.

One point that received much discussion was the need or otherwise to run the analysis of the samples taken for toxicokinetics under GLP conditions with the ensuing validation of the methods, particularly as up until this point, many PK studies in animals were not subject to GLP. These tended to be done in an earlier phase of drug development where assays may not have been perfected or validated. However, it was generally agreed that it would be inconsistent to run pivotal safety studies to GLP standards and not the equally important exposure analysis, and thus, this recommendation became part of the guidance for the first time. It was also recognised that there may be early development PK studies which would not be part of pivotal safety studies, and in these cases, the bioanalysis should be 'fit for purpose' and would not need to be done to this level of stringency.

7.3 What to Measure: Parent/Metabolites/Isomers

As stated previously, when the toxicokinetic guidelines were first introduced, the main objective was to demonstrate that the dosed compound was adequately absorbed and with sufficient bioavailability to provide the required safety margins at the 'no observable adverse effect level' (NOAEL) based on systemic exposure, thereby providing reasonable assurance of the safety of the compound when dosed to humans. Prior to these guidelines, it was not uncommon for a compound to show excellent safety when administered by the oral route, only to find there had been inadequate exposures in the toxicology testing.

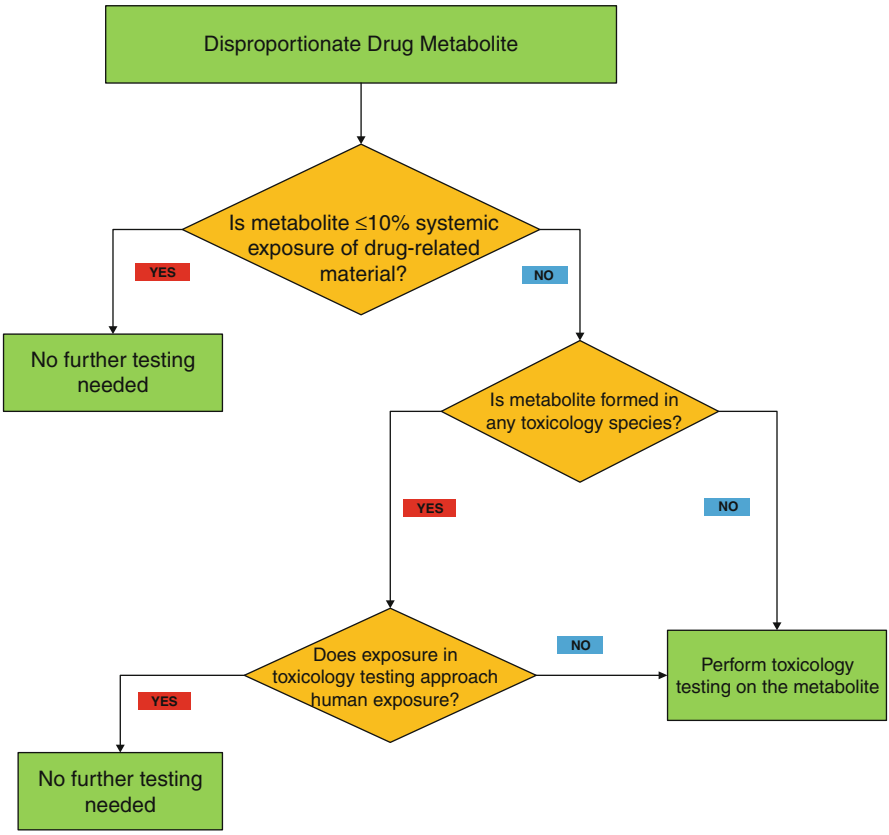
Measurement of the parent compound in the toxicology studies as part of a toxicokinetic evaluation resolved this immediate issue, but it was also noted that there was a need to consider the measurement of metabolites in addition to the parent drug. Within the ICH S3 document, certain circumstances were prescribed where measurements of metabolites should be considered, namely (a) if the parent was a prodrug, (b) when the compound is metabolised to one or more pharmacologically or toxicologically active compounds or (c) when extensive metabolism occurs, and for practical reasons, the measurement of major metabolite is the only practical way of assessing exposure. It was realised at the time that the need for metabolite measurement was

relatively new in safety studies, but it was not clear to what extent this should be done, which metabolites to measure or indeed what to do with the results. On hindsight, this could be thought of as an oversight, but at the time, there was no common view on this issue, and after much discussion, it was decided not to be more prescriptive. At the time of the original guidance, the general view was that with the large multiples of dose used in toxicology compared to the clinical dose, exposure to the metabolites would in any case be covered. It was recognised that many toxic metabolites exhibited their toxicity through reactive metabolites, which by their very nature are short lived and would be unlikely to be measured in biological matrices. In addition, they can be further metabolised by conjugation, such as with glutathione, which are stable, and reduce the metabolite toxicity (Note 10 of the guidance refers to this problem). Thus, because the guidance could be interpreted in different ways and no specific guidance was given on when or which metabolites should be measured, there was some variability in the extent of data provided to the regulatory agencies.

However, since the writing of the original guidance, it has become clear that the situation regarding metabolites was inadequate and more detail was required. The publication of a white paper on the measurement of metabolites in safety testing (MIST) (Baillie et al. 2002) has stimulated regulators to issue new guidelines, often referred to as the MIST guidelines. These state that not only the toxicokinetics of the parent compound should be established but also those of the major human metabolites (Guidance for Industry Safety Testing of Drug Metabolites, FDA 2008; FDA Guidance for Industry 2010a). The original FDA guidance stated that a metabolite should be monitored in toxicokinetic studies if it was ≥10% of the parent compound in human, but this placed an undue burden on the drug development process, especially when a compound is extensively metabolised and metabolites are elimination rate limited. In such cases, the systemic exposures to parent compound will be low with the potential for several metabolites exceeding 10% of parent. ICH M3(R2) recognised this issue and proposed a criterion of ≥10% of drug-related material which was subsequently adopted for the more recent guidelines of both the USA and Europe (Griffini et al. 2010).

These guidances address the processes required to establish that the exposure of all major human metabolites, now defined as ≥10% of drug-related material measured by AUC in human, is adequately covered in the pivotal toxicology studies. A metabolite which is ≥10% of the drug-related material in human and is not adequately covered (see M3 Q&A) in the toxicological species is referred to as a ‘disproportionate metabolite’, and unless there is a reason to the contrary, a different species or a stand-alone safety study with this metabolite would be needed to qualify its safety. The identification of human metabolites can be performed initially using in vitro techniques such as incubations with human microsomes (primarily phase 1 metabolites) and human hepatocytes (phase 1 and 2 metabolites). This is often routinely performed at an early stage when selecting the toxicology species for the first GLP toxicology studies prior to the first-in-human (FIH) studies. Using these in vitro incubations, the human metabolism can then be compared qualitatively with the metabolism in various toxicology species to ensure that there is no metabolite that is unique to human. Again, if a unique metabolite is found, it would be necessary to move to an alternative species or to identify and synthesise the

metabolite and perform a separate toxicological evaluation, all of which will add considerably to the cost and time of development. The FDA MIST guidance is summarised within the original 2008 document by a simple decision tree to help the investigator to decide on the most appropriate course of action, although it has recently been updated in a more recent guidance (FDA Guidance for Industry 2010a) as shown in Fig. 7.1.



A decision tree for disproportionate metabolites

In vitro studies provide an initial direction as to the best species for toxicology testing by comparing the overall metabolite profiles and can even show if there is extensive formation of a specific metabolite. However, they do not give any direct information on the relative plasma exposure of a metabolite in the respective species to the parent, since this is a function of the metabolite's formation, distribution and elimination and the latter two are not known from these studies. To establish exposure, an in vivo evaluation is needed, normally determined by area under the curve (AUC) in human and the animal species used for toxicology. Since this requires

in vivo clinical samples, metabolite exposure measurements are often unavailable for the initial toxicology studies, although the in vitro study may indicate such an extensive formation of human metabolite(s) that the risk of the metabolite exposure exceeding 10% of drug-related material is obviously high.

As a consequence, some investigators might be tempted to perform these studies under steady-state conditions using nonradioactive material to get to steady state, slipping in a radioactive dose when steady state has been reached. Whilst the percentage recovery of that radioactive dose can be calculated, any concentration determinations will be erroneous since the specific activity is unknown with the radioactivity being diluted by all the nonradioactive products formed from previous nonradioactive doses remaining in the body. Thus, although this design can determine if the exposure of a metabolite is greater than 10% of drug-related material for a radioactive dose administered at steady state, it provides no information on the concentration of the metabolite at steady state for safety margin calculation. An alternative is to give repeated radioactive doses, each of a lower amount of radioactivity so the total amount approximates that which would be administered as a single dose. Although the specific activity will be known and constant for all doses, this approach is fraught with practical difficulties, especially if the products have relatively short half-lives compared to the dosing interval since the amount of radioactivity given with each dose will be much smaller and accurate quantitation can suffer. The simplest method, if there is no evidence of dose- or time-dependent pharmacokinetics, is still to use a single radioactive dose design and to use AUC_{∞} as the measure of exposure at steady state, since when at steady state, the AUC_{∞} of a single dose is equal to the AUC for one dosing interval (AUC_{τ}) at steady state. However, if a disproportionate metabolite is found, then for completeness, synthesis and repeat dosing using cold methods would be required (see ICH M3 Q&A, as this is discussed).

One of the issues coming from the MIST guidance is how is it known if all the metabolites have been identified in humans early in development when determination of exposure levels of parent and metabolites are required. But radioactive studies may not be possible. Delaying identification studies to later development can be risky, since if a disproportionate metabolite is found, new toxicology studies may be required, either using a different species or independently evaluating the toxicology of the disproportionate metabolite, seriously delaying the registration and marketing of the drug.

Thus, it could be argued that the drug developer needs to establish the in vivo human metabolism much earlier in drug development at the first-in-human or proof of concept so that disproportionate metabolites can be identified as early as possible and the toxicokinetic programme designed appropriately. Some investigators routinely incorporate a metabolite identification arm into the first-in-human studies, utilising non-radiochemical methods such as LC–MS/MS of plasma, urine and faecal samples. These analyses differ from standard bioanalytical assays by using long chromatography run times to optimise separation of metabolites and full MS scans for identification. Although a good indicator of human metabolism, they are generally not definitive since without a mass/balance, it is impossible to ensure that all of the metabolites are accounted for. Also, it is unlikely that there are reference standards for the putative metabolites at the stage of development when these studies are

performed, so any attempt at quantitation will be by relative response, usually to the parent compound, which may be different to that of the metabolite in question.

Radioactivity remains the best means of accounting for all administered drug-related material, and it is still possible to bring forward such studies to the early clinical phase I development, including the first-in-human studies, providing the amount of radioactivity dosed is so low that it falls within that natural radiation variability. Obviously, the low levels of radioactivity require extremely sensitive instruments for their detection and accurate quantitation, requiring orders of magnitude more sensitivity than conventional liquid scintillation counting. This can be achieved using microtracer studies, sometimes referred to as phase 0 studies (Garner and Lappin 2006) using accelerator mass spectrometry (AMS) and allowing the detection and quantitation in the sub-attomole region. It is beyond the scope of this chapter to go into the detail of this technology, but for the interested reader, there are overviews by Ings (2009) and a series of articles devoted to the subject also in Bioanalysis (2010) Vol. 2, No. 3. However, since AMS does not provide any information on structure, the structural identity of metabolites still uses conventional mass spectrometry.

The MIST guidelines, however, specifically refer to the relative metabolite concentrations at steady state in human at the highest clinical dose but not how it should be undertaken. One question that has not been addressed in the MIST or in the original ICH guidance S3A document (ICH, 1995b) is what to do with chiral drugs. Although it was originally discussed during the drawing up of the S3 document in the context of metabolism, again, detailed guidance was not deemed necessary at the time. The chiral guidance states that when a racemate is developed, the enantiomers should be monitored separately in safety studies and compared to human exposure, but less is discussed with regard to the development of a single enantiomer. Certainly investigation of the possibility of chiral inversion from one enantiomer is required, and if a racemic mixture is developed, then the monitoring of each enantiomer is needed in safety studies unless there is good reason not to do so. However, none of the guidances talk about the monitoring of chiral metabolites in safety studies particularly as the possibility of chiral inversion of not only parent but also metabolites may be more widespread than originally thought (Wsol et al. 2004), and it is well known that there are potentially large species differences in the elimination of different isomers of chiral drugs be they parent or metabolite (Campbell 1990). Another aspect not often considered is if the original compound is achiral, but by the introduction of another moiety through metabolism, a chiral metabolite is produced. Logic says that in such circumstance, each individual isomer needs to be identified, tested and, where necessary, quantified.

7.4 Analytical Methods and GLP

As pointed out earlier, it was decided that analysis for TK associated with pivotal safety studies must be done to GLP with the necessary validations, but as a concession, those not directly related to the interpretation of these studies and particularly

those PK studies undertaken in early development need not be done to this standard. However, this has led to some confusion and to the issuance of 483 noncompliance forms from the FDA since it is difficult to understand when an animal PK study not in some way related to the interpretation of safety studies. Since this is a grey area, many companies now do most of their animal kinetic studies to GLP. It can be argued that in early research, before a drug becomes a development candidate, then it would still be inappropriate to undertake such studies to stringent GLP standards. Little discussion was given in the original guidance document on analytical methods as it was agreed it was not part of the original remit. However, much has been discussed on this topic since then, particularly in relation to their validation with the variety of bioanalytical methods available for the specific and accurate determination of parent drug and metabolites in toxicokinetic studies (Guidance for Industry Bioanalytical Method Validation, FDA 2001).

Whilst there is a plethora of bioanalytical methods available for toxicokinetic analysis, the majority fall into two groups: LC–MS/MS for small synthetic molecules and small peptides and ELISA for macromolecules such as monoclonal antibodies and larger peptides. This is probably now self-perpetuating because most bioanalytical CROs have concentrated their expertise in these two broad technologies, but it should be remembered they might not always be the best methods for some compounds. For instance, very lipophilic molecules lacking a nitrogen atom tend to respond poorly with LC–MS/MS, whereas they might respond extremely well with GC–MS although expertise and equipment in this area are sadly lacking.

7.4.1 Small-Molecule Bioanalysis: Dried Blood Spot Analysis

In the original guidance (ICH, 1995b), it was suggested that limited sampling techniques (composite and population approaches) could reduce the number of animals used to protect the animals from oversampling although it could also reduce the precision of the TK analysis. However, in the intervening years, new techniques have been developed which may overcome this problem, enabling full TK to be done on individual animals (rats). LC–MS/MS has become the bioanalytical method of choice for small molecules. Both chromatography and mass spectrometry are used to provide the selectivity, although most of the selectivity comes from the mass spectrometer with the chromatography separating the analyte of interest from other components of the matrix that can interfere with the response. This allows for short retention and run times, and importantly it can allow for metabolites to be measured at the same time. The increased sensitivity from LC–MS/MS (LLOQ of ≤1 ng/ml) and the relatively high concentrations found in TK samples allows much less blood to be analysed, and this has led to the development of dried blood spot (DBS) analysis.

DBS analysis is a method where small (~20 μl) whole blood samples are placed on a suitable absorbing medium, dried and stored prior to analysis This is ideal for a small rodent species such as a mouse or rat since it enables serial samples to be taken from each animal, even the main study animals, significantly reducing the

number of animals needed for a toxicokinetic evaluation compared to destructive sampling particularly in the mouse (one of the original aims of the harmonization process). Also, as the analytes tend to be more stable when dry, storage can often be at ambient temperature, although this would need to be rigorously tested for each individual analyte, since it will be compound specific. If the increased stability is proven, the DBS samples can be sealed in plastic freezer bags with desiccant and, since no freezing is required, can be transported at ambient temperatures.

Two general procedures appear to be currently used. The first involves collecting the sample onto a suitable absorptive surface which is then allowed to dry fully. It is generally recommended that 2–3 replicate small spots of each sample are collected rather than one larger spot. A disk of known reproducible size is punched out from the centre of the dried spot and eluted with a suitable solvent, usually containing the internal standard. This process can be either semi- or fully automated to 96 well plates (Li et al. 2011). The eluant is then analysed by LC–MS/MS in the normal manner (Wong et al. 2010). However, as the name states, with dried spot analysis, the matrix containing the analyte is blood and not plasma as is traditionally used in toxicokinetic studies. Therefore, if these data are to be related to plasma data, such as those obtained from clinical studies, a correction factor, determined from measuring the whole blood to plasma ratio for each analyte in vitro over the concentration range of the toxicokinetic samples, needs to be incorporated. If the ratio is the same over the whole toxicokinetic concentration range, the correction from whole blood to plasma is a simple constant, but if there is a non-linear relationship, some modelling or in vivo bridging studies may be required. This can be a particular problem for those drugs which are highly concentrated in red blood cells. The more exciting and innovative method with dried spot analysis is a direct desorption procedure without prior extraction of the dried spot, such as desorption electrospray ionisation (DESI) developed by Cooks et al. (2006) and direct analysis in real time (DART) developed by Cody and Laramee (2005). For the former, the DBS cards are cut into strips and mounted securely onto glass microscope slides, whereas for the latter, they are spotted directly onto the glass slide for analysis. The slides are moved through the respective system at a constant rate, allowing the ionspray in the DESI system or heated ionisation gas (helium) in the DART system to desorb the analytes directly from the surface of the dried spot into the mass spectrometer. This has obvious advantages of increasing speed and sample throughput since there is no extraction or chromatography. However, it can be subject to poor sensitivity resulting from ion suppression and possible interference if unstable metabolites are present in the sample. Crawford et al. (2011) compared the pharmacokinetics of different compounds using DBS analysis with DART and the traditional LC–MS/MS method and showed that there was a reasonable correspondence of the data generated by the two methods, although the sensitivity was lower for the DBS method perhaps due to ion suppression. This was especially troublesome if there was no stable labelled internal standard to compensate for ion suppression or if unstable metabolites were present.

However, there are several issues with regard to its general use, namely, acceptance by regulators and the cost. Certain pharmaceutical companies have found that the FDA has asked that if DBS is undertaken for pivotal safety studies, then this

technique must be continued throughout development into clinical PK and that both the dry and wet methods need to be validated completely in all samples, thereby duplicating analysis and cost. The issue here is that the in vivo RBC/plasma ratio of drug can show variability over time, between species, between people and even within the same subject, but whether these differences are so important remains to be seen. A European Bioanalysis Forum consortium of pharma companies and CROs is presently setting out to validate the DBS method to look at the distribution ratio, internal standards, stability and dilution and hopefully within a year will be reporting back to help pharma companies with their interaction with regulators. The EMEA and the MHRA seem to be more pragmatic and are encouraging its use (Beharry 2010). The other issue is one of cost, and the cards are at the moment expensive and are probably only cost-effective particularly in clinical trials when samples from multiple sites across the world need to be transported to centralised analytical labs.

Thus, despite the advantages of collecting small sample volumes requiring fewer animals, especially for the small rodents, together with higher throughput, it is perhaps premature to use these direct desorption techniques for toxicokinetics of formal GLP toxicology studies before more guidance is forthcoming. It is the views of the authors that for the moment, unless considerable validation is undertaken, the potential risk of the loss of precious samples or the inability to interpret the results does not outweigh the potential benefit of using DBS. If S3 were to be written today, it would certainly encourage companies to examine these and other techniques wherever possible so that a database of utility can be drawn up to reduce the number of animals but with the caveats being used on a case-by-case basis and with the necessary validations incorporated.

7.4.2 Large-Molecule Bioanalysis

In the original S3A guidance, there was no consideration to biotechnology-derived pharmaceuticals. However, as more large molecule drugs are being developed, guidance on the need and interpretation of their toxicokinetics has become necessary. In the ICH guidance for biotech medicines S6, it states that pharmacokinetics and toxicokinetics after single and repeated administration should be done wherever possible and that tissue distribution studies may be useful but not mass balance studies. Particular emphasis was placed on ensuring that the methods used in safety studies and those in the clinic were the same or comparable. Also, the potential for species differences in the clearance of these drugs needed to be considered if it relies on the agent being active in the species under test, which often is not the case. Further clarification was provided in the ICH S6 addendum 2009, which tackled the potentially large differences in species activity which can occur for biotechnological products. In this more recent guidance, it suggests that wherever possible, pharmacokinetic (toxicokinetic)/pharmacodynamic relationships should be used to assist in the high-dose safety study selection using, wherever possible, species differences in target binding and in vitro

pharmacology to make the necessary corrections. Thus, it was suggested that the highest safety dose to provide sufficient exposure cover in clinical studies should either be tenfold the maximal effect or, where pharmacodynamics was not measured, tenfold the likely toxicokinetic exposure in humans, whichever is higher.

As stated above, macromolecules, especially the monoclonal antibodies, tend to be highly selective and very specifically interact with the desired target. Also, because of immunogenicity issues, these antibodies are invariably human or humanised which can have profound implications on their toxicology. Unlike their small molecule counterparts, there is rarely off-target pharmacology, so the toxicity tends to be due to the exaggerated pharmacological response from interaction with the therapeutic target. Also, these molecules are typically eliminated by different mechanisms to small molecules, with the main route of elimination being binding to the receptor, internalisation and degradation within the cell. Since these molecules are highly selective, the receptor binding is generally very species specific. Moreover, if the molecule is sufficiently foreign to the host, antidrug antibodies can be produced, which, if neutralising, effectively remove the drug from the system. This will be seen especially when the compound is administered to a species different from that for which the drug is intended, as with toxicology studies. Thus, the selection of species for toxicology studies becomes critical since the species should show cross reactivity with the corresponding target receptor of that species. Small amounts of the drug may also be removed by 'clipping' (metabolism) which can alter the affinity to the receptor or immobilised capture reagent, possibly further impacting the bioanalysis. All these factors play an important role in the development of the typical ligand binding bioanalytical methods (e.g. ELISA) most commonly used for macromolecules.

The physicochemical properties of large molecules, such as monoclonal antibodies, dictate that they will have limited tissue distribution with a relatively small volume of distribution. Also, their clearance tends to be low since the principal clearance mechanism, as stated above, is by binding to the receptor and internalisation. Thus, unlike small molecules which can have therapeutic concentrations in the ng/mL range, these macromolecules tend to be circulating in the μg/mL range, and highly sensitive bioanalytical methods are less critical (Ezan and Bitsch 2009). As mentioned previously, it is important to select a species for toxicology where there is cross reactivity with the human receptor, since the most likely toxicity is likely to be a result of an exaggerated response and it is the main elimination mechanism for these types of compound. Thus, for human monoclonal antibodies, monkey is often the species of choice. If no species shows cross reactivity due to extreme selectivity, it may be necessary to develop an equivalent monoclonal antibody for the toxicology species and examine the toxicology and toxicokinetics of that as a surrogate for the human antibody.

As with small molecules, it is generally accepted that only the free drug is available to bind to the receptor which leads to a further complication of the interpretation of their toxicokinetics. With small molecules, usually the total drug is measured and a correction made for plasma protein binding using data from separate in vitro studies, specifically designed to measure the plasma protein binding over a relevant concentration range. The question is 'what does an ELISA method measure with

macromolecules?' In theory, it should only be the free drug since only the free drug should be available to bind to the immobilised capture reagent. However, with the relatively long incubation times needed for some assays, it is not inconceivable that some dissociation and association occurs with antidrug antibodies which are invariably produced in the toxicology species (Ezan and Bitsch 2009). Thus, the formation of antidrug antibodies needs to be measured for a full interpretation of the toxicokinetic data or, alternatively, if this is not feasible, the measurement of total (obtained by acid pretreatment) as well as 'free' compound, bearing in mind the limitations described previously. However, the formation of antidrug antibodies is not such a simple process since some can be neutralising and others non-neutralising, adding a whole new complexity to understanding a macromolecule's toxicokinetics. Their detection requires different assays with the determination of total antidrug antibodies followed by a more specific assay, normally requiring some form of functional assay to measure neutralising antidrug antibodies.

Compared to the bioanalytical methods for small molecules, the method development times for macromolecules tend to be considerably longer due to the time needed to generate the reagents for an ELISA. This can impact and delay the early development of such molecules since often a suitable bioanalytical method is unavailable for initial pharmacokinetic studies. Workers in this field have, therefore, been looking for alternative methods that can be developed and applied much more rapidly than conventional ELISA methods. With the advent of triple quadrupole mass spectrometers with extended mass ranges, LC–MS/MS now offers a viable alternative, and assay methods based on LC–MS/MS have been reported for some of the lower-molecular-weight proteins and peptides (Kippen et al. 1997; Ji et al. 2003; Buscher et al. 2007). These types of assay, however, are still more of the exception rather than the rule, and a more common approach is to perform a proteolytic hydrolysis of the molecule of interest and to follow the resultant peptides that can act as a 'signature' to the original protein. This was adopted by Barr et al. (1996) and Geber et al. (2003) using stable labelled peptides as internal standard combined with an isotope dilution analysis. These analyses, however, are still in their infancy and with many challenges such as the choice of the 'signature' peptides and extraction of the molecule of interest from a huge amount of irrelevant protein also present in the plasma matrix (Ezan and Bitsch 2009). As a consequence, whilst they may prove useful in filling the gap before ELISAs become available in the Discovery Phase, there is still considerably more work needed to get them to a stage where they can be validated for routine GLP toxicokinetic studies. Clearly we must await advances in the technology and knowledge to catch up with the generalised guidance whilst in the meantime, pragmatic interpretation of the guidances is necessary by regulatory authorities.

7.4.3 Validation of Bioanalytical Methods

Although the need for GLP analytical methods for GLP safety studies was generally accepted during the original discussions on S3, there was very little discussion on

the validation of methods to be used. In more recent years, this has been rectified with numerous papers and guidances from regulators (FDA Guidance for Industry, Bioanalytical Method Validation 2001; Guideline on bioanalytical method validation, European Medicines Agency 2011) and industry-related conferences (Shah et al. 1992, 2000; Shah 2007), all with common features.

The objective of these conferences was to provide a consistent process for establishing accuracy, precision, selectivity, sensitivity, reproducibility, limit of quantification and stability of bioanalytical methods used for drug safety and clinical studies. The Crystal City conference in 1990 examined each of these criteria and how they may be determined, together with acceptance criteria for a bioanalytical run. The second conference focused on the advances made in bioanalytical analyses during the 10 years between the two respective meetings and how they impacted on the bioanalytical process. There was also focus on ligand binding assays, especially with respect to selectivity with possible interference from substances physiologically similar to the analyte and interferences from matrix components that are unrelated to the analyte (matrix effects). A summary of the FDA and European guidelines is given in Table 7.1, but it is recommended that for the interested reader, they should consult the respective guideline documents. Clearly the acceptance of the need for GLP analysis and the ensuing validation in pivotal safety studies has greatly increased the workload of analytical laboratories, but it is felt that this is necessary since the data obtained is essential in calculating human safety margins and starting doses and has to be thought of as an added value.

7.5 Sampling Techniques and Time Points

In 3.3 and 3.4 of the guidance S3 document, there is a discussion on the number of animals to be used, the type of methods and justification for the number of time points. In most cases, this is self-explanatory, but the emphasis of the working party when drafting the guidance was to suggest that sufficient samples should be taken to adequately define the TK profile and subsequent exposure without using too many animals nor to put them under undue physiological stress by the sampling. More recently, the use of composite sampling techniques where a smaller number of samples are taken from any one animal but the complete blood level time curve is taken from all the animals in the group has been proposed and has the advantage of using fewer animals and subjecting them to fewer bloodletting insults, especially in combination with the DBS method (Kurawattimath et al. 2012). The disadvantage, however, is that it complicates the toxicokinetic analysis as is discussed in Sect. 7.11. Numerous different rodent sampling techniques are available and favoured by individual groups ranging from catheterisation, tail vein puncture, tail snip, cardiac puncture, and decapitation (mice) to sublingual and retro-orbital bleeding, which interestingly stress the animals probably less than the animal technician (Grouzmann et al. 2003). Of these techniques, the tail snip is probably the least

Table 7.1 A summary of the FDA and European guidelines for bioanalytical method validation

Criterion	Activity and acceptance criteria
Selectivity	– Interferences due to endogenous substances in the matrix should be tested using six individual and independent sources of the matrix, ensuring selectivity at LLOQ – Check for interference from degradation products in sample work-up, possible co-administered medicines and back conversion of metabolites to parent compound, e.g. unstable N-oxides and acyl glucuronides
Carry over	– Specified in the European guidelines as <20% of the LLOQ for the analyte and <5% for the internal standard as determined by injecting a blank sample immediately after a high-concentration sample (e.g. highest QC or calibration standard)
LLOQ	– The lowest concentration of analyte that can be reliably determined with acceptable accuracy and precision (≤20%) – It should be the lowest concentration on the calibration curve
Calibration curve	– The relationship between instrument response and the concentration of analyte – To include a matrix blank (no analyte and no internal standard), a zero concentration (no analyte but with internal standard) and at least six analyte concentrations over the anticipated concentration range with the LLOQ the lowest and ULOQ the highest – More concentrations to be considered for ligand binding assays. The simplest model should be used for fitting the calibration curve – At least 75% of the back-calculated calibration standards should be within ±15% of the nominal value with the exception of the LLOQ which should be within ±20%
Accuracy—within run	– To be determined using a minimum of 5 determinations at either 3 (FDA) or 4 (European) concentrations – The mean value to be between ±15% except at the LLOQ where it should be ±20%
Accuracy—between run	– Between-run accuracy, must be within ±15% of the nominal value except for the LLOQ which should be within ±20%. The LLOQ, for low, medium and high QC samples from at least 3 runs, 2 of which must be completed on separate days
Precision—within run	– To be determined using a minimum of 5 determinations at either 3 (FDA) or 4 (European) concentrations including LLOQ, low, medium and high QC – The CV should be between ±15% except at the LLOQ where it should be ±20%
Precision—between run	– For the between-run precision, the LLOQ, low, medium and high QC samples from at least 3 runs, 2 of which must be completed on separate days, must have a CV of ±15% except for the LLOQ which should be within ±20%
Dilution integrity	– If the concentration within a sample is above the ULOQ of the calibration curve, it should be diluted to a concentration within the range of the calibration curve – It must be established that the accuracy and precision of the original sample are within ±15% for each of the dilutions used

(continued)

Table 7.1 (continued)

Criterion	Activity and acceptance criteria
Recovery	– The recovery reflects the extraction efficiency of the analyte, and it does not have to be 100% but consistent, precise and reproducible – According to the FDA, recovery should be performed at low, medium and high concentrations, comparing extracted samples with unextracted standards representing 100% recovery
Matrix effect	– In addition to the matrix effects described under selectivity, the European guidelines recommend separately studying matrix effects if an intravenous formulation is used that can cause quenching of response such as polyethylene glycol or polysorbate with mass spectrometric assays – If samples from special populations are used (e.g. pregnant animals), the influence of possible different matrix effects should be evaluated
Stability—short term	– 3 aliquots of the low- and high-concentration QC standards should be thawed as per the assay method and maintained at room temperature for 4–24 h, depending on the assay method and the stability determined – The stability to be determined following 3 freeze/thaw cycles and the stability of each stock solution determined for the period of its intended use – The concentrations in the stored samples should be compared with those of equivalent freshly prepared samples and should be within ±15%
Stability—long term	– The stability of the samples under the condition of their long-term storage should be determined for a period sufficient to cover the anticipated period of their storage – At least 3 aliquots of the low- and high-concentration QC samples should be used – The long-term storage samples should be analysed on at least three separate occasions during the storage period and compared to the mean of the back-calculated values of the respective standards on the first day of the long-term stability testing
Stability—post sample preparation	– The stability of processed samples should be determined to encompass the run time, batch size and conditions of a typical run
Analytical run	– In addition to the study samples, an analytical run should include a blank (no analyte and no internal standard), a zero sample (no analyte but with internal standard), a minimum of 6 calibration curve standards and 3 levels of QC standards (low at ≤3 times the LLOQ, medium at ~50% calibration range and high at ≥75% calibration range) run in duplicate – A run can be accepted if ≥75%, with a minimum of 6 of the back-calculated calibration standards, is ±15% of their nominal value (±20% at LLOQ) – If the calibration standard at the LLOQ fails, the LLOQ should be redefined. Also, ≥67% of the QC values should be within ±15% of their nominal value with at least one at each concentration meeting this criterion. In a multi-analyte assay, this applies to each of the analytes

(continued)

Table 7.1 (continued)

Criterion	Activity and acceptance criteria
Incurred sample reanalysis	– For a total of 10% of samples for studies involving up to 1,000 samples or 5% for studies with >1,000 samples, the samples should be reanalyzed on a separate occasion (day) for each new matrix or for all pivotal bioequivalence studies – A new matrix is defined as plasma, blood or urine or, for example, plasma from different species, plasma from a distinctly different population (e.g. pregnant animals or special patient groups) – The concentration found for the repeat analysis must be within ±20% of the original for at least 67% of the repeats for the bioanalytical of the study to be accepted – If the ISR fails, the cause of the failure should be investigated, and if the bioanalytical method is found to be flawed or lack robustness, the bioanalytical method may have to be redeveloped and the analysis or study repeated

recommended due to the potential for extravascular fluid mixing with the blood during manipulation. Catheterisation and tail vein puncture probably offer the most robust methods, especially for serial bleeding, but are technically more challenging. More recently, whole blood microsampling techniques (WBMT) are being reinvestigated using haematocrit glass tubes enabling serial sampling from the same animals possibly in combination with the dried spot analytical method (Smith et al. 2011). Tubes can be spun to separate the red blood cells and plasma removed by suction or by breaking the tubes in half. These issues, of course, do not extend to sampling from larger animals where concomitant sampling of animals in the main study by venepuncture or catheter is routine.

7.5.1 What to Sample: When Is Plasma/Blood Not Appropriate

Very little comment was made within the guidance on which biological matrix should be measured, stating that plasma, serum or whole blood is normally the choice for toxicokinetic studies, but this would seem to be rather obvious. However, in the M3(R2) guidance, it mentions that when plasma cannot be measured or when high levels are excreted in urine, faeces or bile, these could be alternatives. Faeces and bile are not really viable, especially for orally administered compounds, since for the former, it would be impossible to differentiate between absorbed and unabsorbed compound and, for the latter, a comparison with humans would prove exceedingly difficult. In reality, for most drugs, the body fluids are not the matrix where toxicity is likely to occur, and their analysis can only act as a surrogate for tissue levels, and although this is acknowledged (Sect. 3.6 S3) as a potential problem, in none of the guidances is it clearly stated that if there is a toxicity in a specific organ, then this should be sampled. This was discussed when the guidance

was drawn up, and it was felt that in most cases, tissue samples could not be obtained easily from humans, and thus, as a general rule, animal tissue analysis should not be recommended since it would add an unnecessary burden on drug development without practical usefulness. This does not preclude determining tissue levels to better understand the mechanism of toxicity. Also, in those specific cases where an understanding of the comparative tissues levels becomes important, exposure comparisons can be made between animals and humans without actually taking biological samples, for example, brain uptake of a fluorinated compound by MRI where toxicokinetic/toxicodynamic comparisons were made with both animals and humans (Campbell 1995) to compute CNS human safety margins. In this case, brain uptake in humans was distinctly different to the animals used in safety assessments and confounded the interpretation of the results. Use of PET ligands is another method that can be used, but with the caveat of not always being able to differentiate between the levels of intact drug or metabolic products. It has the advantage of being used early in investigatory microdose phase 0 studies. Finally, there is specific guidance in S3A with regard to the need for embryo, foetal or newborn exposure but, when it is needed, is rather ambiguous and may not be routinely undertaken (see Sect. 7.8.4).

7.6 Exposure Quantitation and Interpretation

7.6.1 Use of TK in Dose Level Selection and Study Design (Frequency)

The determination of the maximum safe starting dose is well documented by the FDA in two guidelines: the first referring to compounds in general (FDA Guidance for Industry: Estimating the maximum safe starting dose in initial clinical trials for therapeutics in adult healthy volunteers 2005) and the second referring specifically to anticancer compounds (FDA Guidance to Industry: S9 Nonclinical evaluation for anticancer pharmaceuticals 2010b). The former document describes how the starting dose for a first-in-human study is calculated, which does not employ pharmacokinetic principals per se, but uses the dose level at the no observable adverse effect level (NOAEL), converting a mg/kg dose level to a human equivalent dose (HED) by normalising to body surface area calculated from body weight$^{0.67}$. For anyone needing to perform this, the calculations have been further simplified in the guidance document with a table of conversion factors, depending on species, needed to obtain the HED also shown in Table 7.2.

The overall process described in the maximum safe dose guidance has been summarised by a decision tree on the last page of the guidance which is also presented in Fig. 7.2.

The guidelines for anticancer agents tend to be more flexible since first-in-human studies are conducted in patients with serious and life-threatening malignancies. These patients are hopeful of a chance for a positive response to their cancers, so it

Table 7.2 Conversion of animal doses to human equivalent doses (HEDs) based on body surface area (FDA Guidance for Industry 2005)

Species	Reference body weight (kg)	Working weight range[a] (kg)	Body surface area (m^2)	To convert dose in mg/kg to dose in mg/m^2 multiply by km[b]	To convert animal dose in mg/kg to HED[c] in mg/kg, either	
					Divide animal dose by	Multiply animal dose by
Human	60	–	1.62	37	–	–
Child[d]	20	–	0.80	25	–	–
Mouse	0.020	0.011–0.034	0.007	3	12.3	0.081
Hamster	0.080	0.047–0.157	0.016	5	7.4	0.135
Rat	0.150	0.080–0.270	0.025	6	6.2	0.162
Ferret	0.300	0.160–0.540	0.043	7	5.3	0.189
Guinea pig	0.400	0.208–0.700	0.05	8	4.6	0.216
Rabbit	1.8	0.9–3.0	0.15	12	3.1	0.324
Dog	10	5–17	0.50	20	1.8	0.541
Primates:						
Monkeys[e]	3	1.4–4.9	0.25	12	3.1	0.324
Marmoset	0.350	0.140–0.720	0.06	6	6.2	0.162
Squirrel monkey	0.600	0.290–0.970	0.09	7	5.3	0.189
Baboon	12	7–23	0.60	20	1.8	0.541
Micro-pig	20	10–33	0.74	27	1.4	0.730
Mini-pig	40	25–64	1.14	35	1.1	0.946

[a]For animal weights within the specified ranges, the HED for a 60 kg human calculated using the standard km value will not vary more than ±20% from the HED calculated using a km value based on the exact animal weight

[b]km is 9.09 × body weight$^{0.35}$

[c]Assumes 60 kg human. For species not listed or for weights outside the standard ranges, human equivalent dose can be calculated from the formula: HED = animal dose in mg/kg × (animal weight in kg/human weight in kg)$^{0.33}$

[d]The km value is provided for reference only since healthy children will rarely be volunteers for phase 1 trials

[e]For example, cynomolgus, rhesus and stumptail

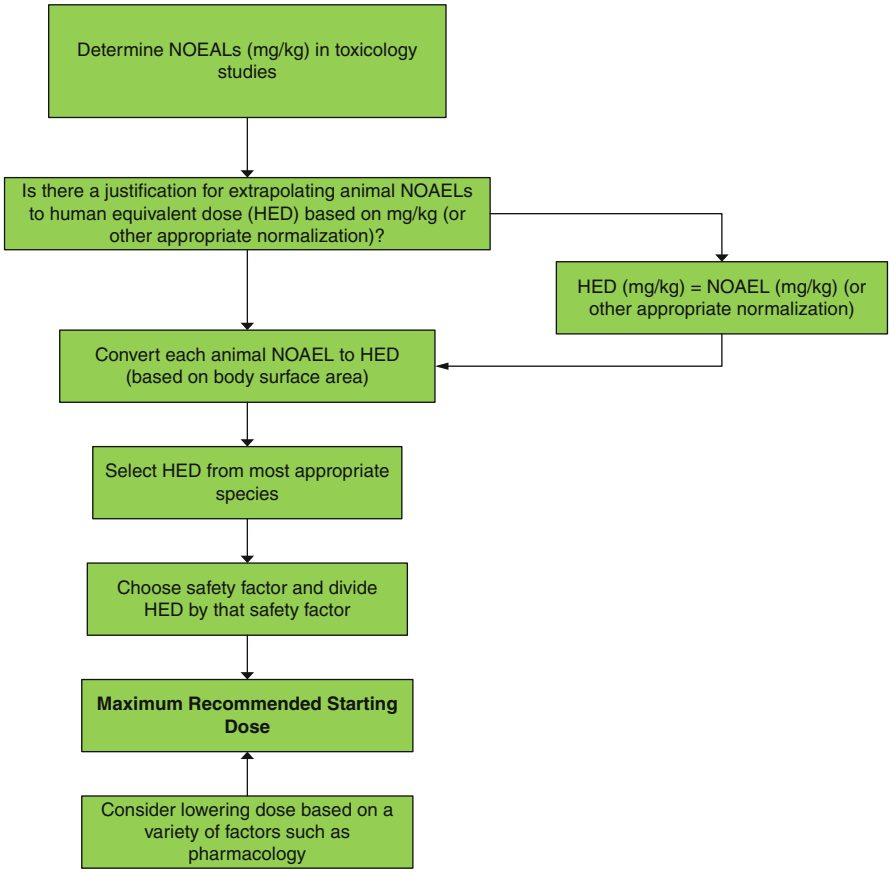


Fig. 7.2 Process for the selection of the maximum safe starting dose in healthy volunteers

is unethical to use dose levels so low that efficacy could not be expected. Thus, for small molecules, a starting dose is commonly selected as 1/10th of the severely toxic dose in 10% of the animals when rodents are used (STD10) or 1/6th of the highest non-severely toxic dose (HNSTD) when non-rodents are used. The choice of species will depend on which is the most sensitive to the toxicity of the compound. For biopharmaceuticals showing immune agonistic activity, a minimally anticipated biological effect level (MABEL) should be considered.

Although these two guidances provide a means of calculating an initial dose for first-in-human studies, neither is based directly on pharmacokinetic/toxicokinetic and exposure measurements which nullifies the use of toxicokinetics. Both use body surface area for extrapolation which is not surprising since clearance extrapolation approximates better to body surface than body weight, but that is the extent that pharmacokinetics is used. However, it must be remembered the maximum safe starting dose guidance (Fig. 7.2) states that a suitable safety factor is needed which should be justified and, where possible, an even further reduction of the dose should

be explored based on various preclinical findings. This is where pharmacokinetic considerations can be applied, but the challenge is the prediction of the human exposure. For toxicology studies conducted after the first-in-human studies, there is no problem, but for those toxicology studies conducted to support the first-in-human study, the human exposure has to be extrapolated from the preclinical data. There are two approaches commonly applied: an in vitro approach typically using human microsomes or hepatocytes and an in vivo approach using allometric scaling.

7.6.2 In Vitro

The in vitro methods all involve the determination of intrinsic clearance usually from microsomal or hepatocyte incubations. Careful thought should be given as to which of these preparations should be used since microsomes primarily determine the intrinsic clearance resulting from phase 1 metabolism (e.g. cytochromal), unless the incubates have been fortified to include some glucuronyl transferase activity. Hepatocytes, on the other hand, measure all enzymes, both phase 1 and phase 2, although one must be reasonably confident that the test compound can actually distribute into the cell. Conceptually, hepatocytes should be the more robust preparation, but if it is known from previous studies that the test compound is exclusively metabolised by cytochromal enzymes, microsomes offer a viable and technically easier alternative.

These data can then be used directly either to predict human plasma clearance and, subsequently, exposure or can be incorporated into physiologically based pharmacokinetic models with permeability and physicochemical data to simulate the human plasma profiles. The two most commonly used software packages are GastroPlus produced by Simulations Plus, Inc. (CA, USA) and SimCyp (Sheffield, UK). Both perform similar functions but SimCyp integrates the physiologically based pharmacokinetic model approach with experimentally validated population distributions of individual cytochromes and transporters. Thus, SimCyp can predict not only the average human pharmacokinetics but also the population pharmacokinetics. This also allows predictions for different patient populations such as different ethnicity, age (young and elderly) and disease state since their population distributions form part of the database. SimCyp is also a very powerful tool for predicting the effect of competitive and mechanism-based drug–drug interactions using in vitro inhibition data of cytochromes and transporters, since it is also capable of predicting extreme situations or combinations of situations where the problems are most likely to arise.

7.6.3 In Vivo

An alternative approach for extrapolating from in vivo animal pharmacokinetic data to that in human is the use of allometry. This approach has been around for centuries,

when Galileo observed that large animals had disproportionally larger bones than those of smaller animals. It was further applied to a variety of physiological parameters and taken up in the 1970s and 1980s for application to pharmacokinetics in drug development by Dedrick (1973) and Boxenbaum (1982). It was also extensively reviewed by Ings (1990) and more recently by Mahmood (2007).

Basically, allometry relies on an empirical mathematical relationship between the physiological or pharmacokinetic parameter of interest and the body weight of the animal using a power regression as shown by (7.1):

$$\text{Pharmacokinetic parameter} = \text{coefficient} \times \text{body weight}^{\text{power}} \tag{7.1}$$

The simple approach described above generally works reasonably well for those compounds eliminated mainly by flow-dependent processes, such as renal clearance or those compounds with a high hepatic clearance approaching hepatic blood flow. However, it tends to fail when there are large species differences in plasma protein binding and for low clearance, extensively metabolised drugs. The former can easily be rectified by correcting for plasma protein binding, but the latter is more complicated since human has developed a slower metabolism than other mammalian species, including metabolism by cytochrome P450. This is correlated with a larger brain weight and longer maximum life span (MLP) for human so that either brain weight or MLP can be empirically incorporated into the allometric equation to correct for the slower human drug metabolism relative to other species. In addition, a major challenge for the prediction of clearance is the assessment of uncertainty around the computed values, and the use of mechanistic physiology may provide a greater confidence (Lavé et al. 2009).

Both the in vitro and in vivo approaches have been developed mainly with small synthetic molecules in mind, but what about macromolecules, including proteins, since the number of these types of molecules entering development is ever increasing. The in vitro methods do not apply as these compounds are not eliminated via cytochromal enzymes, and thus, with the state of our current knowledge, allometry appears to be the only viable alternative. However, its relevance will still depend on the primary eliminating system. If the eliminating system is common to all of the species used for allometry, then the standard allometric approach using different species will probably be appropriate (Khor et al. 2000). If, however, the macromolecule is a monoclonal antibody, which is normally eliminated by binding to the receptor and internalisation, only the species with cross reactivity to the target receptor can be used, and this is typically restricted to a primate. This invariably limits the scaling to single species, so the scaling has to be performed using a standard power function (e.g. 0.75 for clearance and 1.0 for volume of distribution). The viability of this approach has been investigated retrospectively, using a variety of monoclonal antibodies that had already been administered to human (Dong et al. 2011), demonstrating that it was appropriate when the pharmacokinetics were linear, but for those monoclonal antibodies showing non-linear pharmacokinetics, it was only predictive once receptor saturation was achieved.

It is the authors' experience that the in vitro methods generally provide the most robust method of predicting human clearance of small molecules, providing that precautions such as using concentrations below the Km and correcting for non-specfic binding are taken into account, such as correcting for non-specific binding in the incubates. Ideally, both in vitro and in vivo approaches should be applied and the predictions between the two compared. If they are similar, it provides some confidence in the predicted values, but if there is a wide variance, the cause may need to be investigated. There are occasions where such a discrepancy cannot be resolved preclinically, and then the best option might be to go directly to human using microdosing employing a very abridged toxicology package as set out in the respective FDA and EMEA guidelines (FDA Guidance for Industry: M3(R2) Nonclinical safety studies for the conduct of human clinical trials and marketing authorization for pharmaceuticals 2010a; ICH Topic M3 (R2) Non-clinical safety studies for the conduct of human clinical trials and marketing authorization for pharmaceuticals 2009). Scaling techniques were again discussed in the working party putting together the S3 document, but at that time, not enough general experience was available to say if this concept really had practical usefulness. It is now felt that a combination of in vitro and in vivo extrapolation techniques together with the HED corrections provides the best framework for assessing a safe starting dose in humans particularly if combined with PK/PD calculations and should be the subject of any future guidance on this subject.

7.7 Non-Peripheral-Absorbed Routes of Administration

There is no guidance as to what to do with regard to TK when the route of administration is different to those most commonly encountered, particularly where the drug is supposed to be non-absorbed into the systemic circulation. What needs to be measured for dermal preparations or those compounds directly administered into the brain or remain in the GI tract such as antibiotics or those treating IBS or colitis? Again a pragmatic approach is required, and normal TK sampling is needed to show the extent of systemic exposure, if any, and where necessary attempt to measure the levels of the drug and metabolite(s) at the site of action. However, it is always difficult to prove a negative and may be necessary to perform additional toxicity studies using a parental route (e.g. intravenous) just to obtain the necessary exposure for identification of the target organ toxicity.

7.8 TK in the Various Areas of Safety Testing: When to Use

7.8.1 In Vitro

In the guidance documents, no consideration was given to in vitro toxicokinetics, which in hindsight was perhaps an omission and was never discussed to any great

extent in the working parties. In none of the guidances is there a suggestion to measure incubation medium during in vitro tests since presumably it was thought that if a concentration was added, that was the concentration present throughout. No thought was given to the solubility, the chemical or metabolic stability of the drug, adsorption onto the apparatus or binding to the matrix, all of which can result in profound differences between the nominal concentration added and that actually present. Genotoxicity studies such as the Ames and in vitro micronucleus test use rat S9 fractions to activate the system and to produce metabolites to test their toxicity, whilst the hERG electrophysiological screen or the Purkinje fibre assay to evaluate for potential of QTc changes uses systems that are unlikely to extensively metabolise the parent. Clearly when rat S9 is co-incubated, it may be expected that metabolism will occur but without measurement of the incubation medium; this is an assumption, not a fact. It also begs the question of why rat S9 is used, especially if there is a clear species difference and where metabolism by human S9 would be more relevant. Certainly studies by Rueff et al. (1996) investigating the potential for other metabolising systems showed that compounds belonging to many different classes of chemicals were activated to genotoxicants, including carcinogens, which were negative in the rat liver S9-mediated assays. More recently, discussion was held at the International Workshop for Genotoxicity Testing, Basel (2009) on the use of different cell lines to minimise false positives. It was concluded that the use of p53-competent cell lines could be used as well as a hepatoma cell line, HepaRG, since these cells possess better phase I and II metabolising potential compared to cell lines commonly used in this area and may overcome the need for the addition of S9 fractions for activation. This finding has also been confirmed in human HepG2 transformants which may be the more relevant system to predict toxicity in humans (Hashizume et al. 2010). Thus, in this evolving story, rat S9 may not be the most appropriate activation system as stated in S2B, Section 1b, of the original S3 guidance but, until consensus is reached on alternative systems, measurement of incubation medium may be the best way of ensuring a meaningful interpretation of results and could be included in any further guidance.

7.8.2 Safety Pharmacology

The S3A guidance has no mention of the need to determine toxicokinetics in safety pharmacology since this is a concept derived from specific ancillary pharmacology. In a subsequent safety pharmacology guidance, ICH S7A, again there was no mention of the need for concomitant toxicokinetic investigations. In practice, if the pharmacokinetics at the doses using similar formulations and dose routes has been established in the same species as used in safety pharmacology, there appears to be little need to do additional toxicokinetics unless the results warrant further investigation. However, if the pharmacokinetics is unavailable or non-linear, then consideration should be given for the need of additional toxicokinetic information. Certainly in the interpretation of the results from such studies, consideration should

be given to the PK–PD relationships of any findings and the potential safety margins for humans based on unbound plasma levels. There is at least one exception to the generalities mentioned above and that is when specific cardiac rhythm changes are noted such as QTc prolongation. This was not included in the original S3 document, but in the S7B guidance on QTc changes, it states that monitoring plasma drug and metabolite plasma levels would be useful in interpreting changes and appropriate models can be conducted to help the design of future studies. On hindsight perhaps 'useful' should be changed to essential since the need to do plasma level monitoring in Thorough QTc Studies in humans is paramount.

7.8.3 Short- and Long-Term Toxicity

One of the most important aspects of the guidance, which was a common feature of the discussions in the working party, was the necessity of measuring exposure whenever possible in safety studies but particularly in those deemed pivotal. Exploratory animal pharmacokinetic analysis should be done particularly in research and candidate selection but not to GLP standards. However, as soon as an analysis is undertaken for safety studies which are used to understand the toxicity of the compound and are well-designed studies performed to GLP standards, the same standards should be applied to the toxicokinetics perhaps irrespective of if they are 'pivotal'. Before this was formalised, it was rather a grey area and groups chose whether or not to do toxicokinetic analysis. Also the emphasis of the guidance was that these results at any stage in development, if done well, would then help design future studies with regard to formulations, frequency of dosing, sampling times and which analytes to measure and for how long. The need to measure levels at the beginning and end of studies as a minimum was also emphasised and has now become standard practice within the industry.

7.8.4 Reproduction and Fertility Studies

When formulating the initial guidance, some discussion took place with regard to how much toxicokinetics should be required in reproduction and fertility studies since there was no universal consensus of what was needed. It was generally accepted that preliminary exposure determinations are required, particularly in the rabbit, to show that the drug is absorbed and to produce a likely exposure for a given dose in a species not ordinarily being used for other safety studies. This was advocated in the FDA Reproduction Guidance S5A. However, it was not deemed necessary to conduct these studies whilst the animal was actually pregnant or indeed in lactating animals since it would be difficult to prescribe at what time in the development cycle such analysis should be done. In the original guidance, it was suggested that toxicokinetics monitoring during the study would not be generally needed,

particularly for drugs with low toxicities (Sect. 7.6.1), but this appears to be a flawed argument since the toxicities would not be known until completion of the studies by which time it would be perhaps too late to undertake toxicokinetics. In practice, satellite animals can be dosed under the same conditions as the main study with sampling at various times during gestation to measure the exposure, particularly important where there is evidence to suggest that pregnancy can alter the toxicokinetics and subsequent toxicity of drugs (Miida et al. 2008). A more recent Reproduction Guidance (FDA Guidance—Reproductive and Developmental Toxicities—Integrated Study Results to access concerns Sept 2011) indeed puts a larger emphasis on the need for detailed plasma exposure data to be compared to that found in humans and using the criteria to define risk: concern ≤10-fold safety margin ratio and acceptance ≥25 ratio. The recent guidance, however, does not mention placental transfer, and whilst the S3 guidance is less than clear on the need to undertake embryo–foetal placental transfer studies, it does state that 'Toxicokinetics may involve exposure assessment of dams, embryos foetuses or newborn' with a note that says that whilst it is important to consider the transfer of substances entering the embryo–foetal compartment, foetal exposure is a parameter most often assessed in separate studies. Notwithstanding, it is the authors' view that placental transfer should be undertaken using satellite studies to enable adequate interpretation of the reproduction studies since if no toxicities are found, it is needed to show that the foetus was, in fact, exposed to the drug and/or metabolites, whilst if toxicities are found, it is still necessary to understand what the concentrations are and what entity may have been the cause. Normally this would be done only in the rat and assume that the same would occur in rabbits and subsequently in humans unless there is good reason (different results) to do some investigatory work in the rabbit as well. There is nothing in the S3A or S3B or other guidances on the need to measure drug concentrations in milk in the reproduction studies to see if large amounts of drug-related material could be ingested by the offspring. It was originally decided not to include such recommendations as a general rule since milk constituency varies considerably among species and the relevance to human milk is questionable. This can be better addressed using physicochemical parameters (pK_a, Log *P*) and in vitro human milk protein binding. Failing that, any particular safety question to address milk transfer could be left to later in development with small studies in lactating women which would provide more relevant data and information.

7.8.5 In Vivo Genotoxicity

The guidance is very succinct and says that if negative results are found with the in vivo tests, such as the rat or mouse bone marrow micronucleus test, then it may be appropriate to have demonstrated systemic exposure or to have characterised exposure in the indicator tissue (e.g. bone). The genotoxicity guidelines (S2B) say much the same thing. Under normal circumstances, the pharmacokinetics will have already been done either as stand-alone experiments or the toxicokinetics

at higher more relevant doses used in safety studies and, thus, does not need to be repeated unless specific organ uptake is needed to assess the exposure of the toxicant.

7.8.6 Carcinogenicity Studies

Guidance on the use of toxicokinetics in carcinogenicity studies is straightforward with no surprises except for the caveat that there may be differences in the pharmacokinetics in the strains which are used, particularly mice, and separate pharmacokinetic studies may be needed prior to starting long-term carcinogenicity studies strain of choice. Certainly when the drug is administered in the diet rather than by gavage or by a different route to that used before, such studies are essential to be sure that sufficient exposure is achieved (Note 13). Samples should be taken from satellite animals at different periods throughout the duration of the study since the levels achieved at the end of 18–24-month dosing can be markedly different to those at the start of dosing. For example, for rats in particular, since they are dosed on a mg/kg basis, male rats gain weight throughout the study, and unless the drug has a large volume of distribution, plasma levels will increase not because of clearance reasons, but just because more drug is administered. Conversely females go through their menopause during the study, and the hormonal changes can produce unpredictable effects on the clearance of drugs. The use of TK in assessing dose selection for carcinogenicity studies was not detailed in the original S3 document but has been well described in ICH S1C(R2) ICH (1995a) particularly for the high dose where it has been stated that a safety margin of at least 25-fold for unbound drug or metabolites based on total exposure (AUC) would be acceptable. However, provisos were listed including the need for the same strain in TK and test species, over a suitable long time; comparison of metabolism in species and human; decision on which is the active moiety; differences in protein binding; and comparison should be made with human PK studies at the maximum recommended daily dose. It was also noted that the rodent TK need not be undertaken during the carcinogenicity study itself and could be taken from separate subchronic studies, thereby saving animals and resource. However, this could lead to problems that if for any reason the in-life carcinogenicity study was compromised, there would be no means of assessing actual exposure in the study.

7.9 TK/TD Analysis: Use of Safety Biomarkers

Although it has long been recognised that blood drug concentrations are only a surrogate for activity (Peck et al. 1994), rarely is there any attempt to undertake toxicokinetic-toxicodynamic analysis even though large amounts of such data are available in safety studies. Increasingly, more specific biomarkers for toxicity

showing tissue specific involvement are available (Muller and Dieterle 2009), and the possibility of undertaking toxicokinetic-toxicodynamic analysis becomes increasingly possible. Inspection of the literature, however, indicates that it is more frequently reported in environmental assessments particularly for metal intoxication and for anticancer drugs to assess the first dose in humans (Batra 1995) and rarely elsewhere. Recently Hoshino-Yoshino et al. (2011) have used toxicokinetic-toxicodynamic assessment to investigate a number of tyrosine kinase inhibitors, and Zhong et al. (2000) have used it for an anxiolytic compound that produced respiratory depression and showed a very steep exposure/response curve, suggesting adverse effects occurred over a small concentration range. It was for this reason the drug development was stopped. In another study, Campbell (1995) undertook toxicokinetic-toxicodynamic analysis of brain levels of fenfluramine and its active metabolite with changes in 5HT brain levels as an index of toxicity in a number of species including non-human primates. These were compared to those measured by MRI in humans at therapeutic doses and showed that serotonin levels would be unchanged. Peterson et al. (2011) has more recently reviewed the use of imaging techniques (NIN) such as MRM, PET, SPECT and ultrasound in small animals to measure of pathological changes by non-invasive techniques. These technologies allow non-terminal, time-course evaluation of functional and morphologic end points and can be used to translate biomarkers between preclinical animal models and human patients which, in turn, can be linked to concentrations in the appropriate biological matrix. It is unlikely that these newer techniques will be used routinely in the near future but should be considered when a specific and critical issue needs to be evaluated further.

Perhaps one of the reasons why all these technologies are not used more often is the plethora of guidances which has been produced on the qualification and validation of biomarkers (Guidance for Industry : pharmacogenomics data submission 2005: Biomarker qualification pilot process at the US food and Drug administration 2007; Biomarker Qualification: Guidance to Applicants 2008a; Final Report on the Pilot Joint EMEA/FDA/VXDS experience on qualification on Nephrotoxicity Biomarkers 2008b; Guidance for Industry Bioanalytical Method Validation 2001; Fit for purpose method development and validation for successful biomarker measurement 2006). In brief, before a biomarker can be accepted as a useful index of activity or toxicity, the method must be fully validated and most importantly qualified by categorically linking it to a preclinical or clinical end point. Recently a consortium of regulators, academics and pharmaceutical scientists (Predictive Safety Testing Consortium—PSTC) has shown that certain nephrotoxicity biomarkers can be used for the non-invasive detection of renal toxicity in preclinical models and could also be used clinically to monitor renal disease progression. The growing use and acceptance of such biomarkers in the future, in conjunction with toxicokinetic-toxicodynamic modelling, will certainly help improve the understanding of the relationship between drug level exposure in animals, their toxicities and the translation to humans for the early detection of problems.

7.10 Tissue Distribution Studies

Although quite separate from classical toxicokinetic studies (ICH designation S3A), ICH also issued guidance on repeated dose tissue distribution studies (ICH designation S3B (ICH, 1994)) at exactly the same time (October 1994) with subsequent adoption by the USA, Europe and Japan. As a consequence, these two guidances tend to be linked. The guidance on repeated dose tissue distribution was developed largely to resolve the ambiguity of when repeated dose tissue distribution studies are needed to support toxicology, since Japanese developers tend to put far greater emphasis on such studies than either the USA or Europe due to the historical unease regarding the use of radioactivity in humans. This guideline describes four circumstances when repeated dose tissue distribution should be considered, the operative word being 'considered'. However, unless there is a very good reason for doing these studies, in practice, they are seldom undertaken. In summary, the reasons given are:

1. When the apparent half-life of the test compound and/or metabolites in tissues and organs exceeds the terminal half-life in plasma and is more than twice the dosing interval in toxicity studies
2. When steady-state concentrations of test compound/metabolites determined in repeated dose toxicokinetic studies are markedly higher than those predicted from single-dose studies
3. When there are histopathological changes critical to the safety evaluations that were not predicted from short-term toxicity studies or other single-dose studies
4. When a drug is being developed for site-specific targeted delivery

Before starting any tissue distribution study, the objective of the study has to be carefully defined, and serious consideration should be given as to whether the data generated will resolve the specific issue under investigation. In addition, the design of the study has to be carefully thought through, such as what should be measured: total drug-related material, parent compound, metabolite(s), total compound or unbound compound? Is it best to perform the study as a single dose or under steady-state conditions either by repeat dosing or constant rate infusion?

There are many designs available, but the most commonly used are those which involve the dosing of radioactive compound followed by either tissue excision and liquid scintillation counting or quantitative whole body autoradiography (QWBA) with phospholuminescence to determine total radioactivity. These traditional tissue distribution studies tend to be poorly predictive of target organ toxicity, since most tissue distribution studies only measure total drug and/or metabolite and it is widely accepted that only the free (unbound) drug is available to interact with receptors, enzymes or other biological molecules. As discussed previously for chronically administered drugs, these studies should be performed at steady state, when the system is in true equilibrium, and the average free unbound concentrations will be the same in the target tissue as in plasma. However, if the system is not at steady state, direct measurements of the free compound would need to be made by techniques such as microdialysis since this measures only free compound. Alternatively,

a tissue of interest can be excised and homogenised and the free fraction in the homogenate determined using either ultrafiltration or equilibrium dialysis. These approaches are rarely performed, although equilibrium dialysis was successfully applied to a variety of marketed drugs with widely differing physicochemical properties and affinity to the efflux transporter, P-gp, to determine the free drug concentration in plasma and brain (Kalvass et al. 2007). Another question often asked is, when should these tissue distribution studies be performed? The S3B guidance (ICH, 1994) suggests that it should be investigated early in drug development and certainly by phase 1. It is the view of the authors that unless it could resolve a specific problem, it need not be done until later, once a proof of concept clinical study has been performed, but before the human radiolabelled study since these data are also needed for radioactive dosimetry.

7.11 PK Analysis, Modelling and Statistics

The ICH S3 guideline for toxicokinetics specifically states that 'The data should allow a representative assessment of the exposure. However, because large intra- and interindividual variation of kinetic parameters may occur and small numbers of animals are involved in generating toxicokinetic data, a high level of precision in terms of statistics is not normally needed'. Thus, most toxicokinetic analyses tend to be limited to descriptive statistics (mean ± SD and/or median with range) of the most important parameters, typically C_{max}, T_{max} and AUC_{0-24h} since C_{max} and AUC are cited as the main parameters of exposure in the ICH S3A guideline. Half-life may also be calculated for the last dose, although it does not generally enter into toxicokinetic considerations. Depending on the duration of the toxicology study, toxicokinetics are normally obtained for the first dose and last dose, and for the longer term studies (e.g. >3 month), a mid-dose evaluation is usually included. However, the guidance did not make any comment on the methodology to use to obtain the TK parameters and has left it to the individual with variable consequences.

Unfortunately, toxicokinetic analyses have tended to become a black-box process with the plasma concentrations plugged into a commonly used validated package such as WinNonlin (Pharsight, CA) or SAS (SAS Institute, NC) where the parameters are generated automatically and tabulated ready for direct inclusion into the report. No doubt, the numbers entered into the programme and transcribed into the report will be rigorously quality controlled both by the scientists performing the analysis and by the Quality Assurance Department, but how much thought goes into the understanding of how the individual parameters are generated and is the process the most appropriate for the compound in question?

Let us take an example of a toxicokinetic sampling scheme, but instead of using measured plasma levels, the plasma levels are predicted from (7.2) for oral dosing without a lag time followed by a biexponential decline. This equation was also used to calculate the true AUC_{0-24h}:

$$\text{Plasma conc} = 9{,}000e^{-2.0t} + 594e^{-0.139t} - 9{,}594e^{2.67t} \tag{7.2}$$

where t is time.

Now let us now assume the toxicokinetic protocol used a typical sampling scheme: predose and 5 time points after dosing. Usually one of these time points will be at 24 h after dosing, since it is the 24 h exposure that is generally calculated. The other four time points tend to be distributed within the working day of when the compound was dosed, based on either data from previous pharmacokinetic or dose range finding studies. Pragmatically, one time point will approximate the C_{max}, and another will be at the end of the working day (e.g. 8 h) with two further points in between. For our example, let us use predose: 0.5, 1, 4, 8 and 24 h. The AUC_{0-24h} was then determined using the five toxicokinetic time points with the simulated plasma level data using the traditional and often default linear trapezoidal rule and by the lesser used but statistically more correct linear trapezoidal rule for the ascending portion and logarithmic trapezoidal for the descending portion of a plasma profile.

The true AUC_{0-24h} for the curve calculated by integration of the exponential equation is 5,028 ng h/ml, but when calculated using the typical time points selected for a toxicokinetic study with the linear trapezoidal rule, a value of 5,869 ng h/ml was obtained, that is, a 17% systematic overestimation. On the other hand, using the linear trapezoidal rule for the ascending and log trapezoidal rule for the descending, a much more accurate AUC_{0-24h} (5,157 ng h/ml) was obtained, giving only a 2.6% systematic overestimation. This example was taken at random, and the systematic errors resulting from the default use of the linear trapezoidal rule can change, depending on the shape of the plasma profile and the frequency of sampling. Past experience has shown that systematic errors >20% can be obtained without too much difficulty (Ings 1989). It seems paradoxical, after spending so much time and resource on reducing experimental and bioanalytical error with validations, GLP audits, QC and QA, that we are still willing to accept a systematic calculation error that can easily be avoided. Unfortunately, the method of calculation of exposure was not described in the original or subsequent guidelines, but is something worth considering for the future, especially when it is so easy to resolve.

All the above approaches use either the average parameters from individual animals or, in some cases, data from pooled animals. One of the challenges, especially with the use of small animal species, such as mouse or rat, is that satellite groups have to be used, so there are no data from the animals actually on the main toxicology group. Thus, when there is an animal(s) showing an adverse effect, there is no way of determining the exposure for that particular individual, since extensive blood sampling from the main toxicology group is generally avoided as it can confound the toxicological findings. However, by a small modification to the toxicokinetic plan, it is possible to overcome this limitation and determine the exposure for the animals in the main toxicology group using mixed effect (population) modelling (Aarons and Graham 2001; Hing et al. 2002; Ingwersen et al. 2002) and by using microsampling and DBS. A possible protocol would still include a

satellite group which will be relatively data rich and will provide the backbone for the modelling. However, at selected times distributed throughout the toxicology study and dosing interval, a few blood samples can be collected from the main study group. The loss of blood volume should be minimal because most bioanalytical methods now only require a few microlitres of blood or plasma and only a very few samples are needed from each animal so the toxicological end points should not be compromised. By combining the sparse data from the main toxicology group with the relatively data-rich satellite group, robust model parameters can be obtained using mixed effect modelling (e.g. NONMEM, University of San Francisco, CA) that will allow prediction of the full pharmacokinetic profile of individual animals on the main toxicology group using Bayesian feedback. In addition, depending on the compound, it may also be possible to dispense with or significantly reduce the number of animals in the satellite group, thereby reducing the total number of animals used for such studies.

As stated at the beginning of this section, the original ICH guidelines state that a high level of statistical precision is not needed, although an average value (mean or median) and some form of indication of the variability (standard deviation or range) are expected. Both can be generated from serial sampling of individual animals or when the mixed effect model approach is used, but it can become more difficult if destructive sampling is employed where the animal is sacrificed after sampling. Bailer's method (Bailer 1988) is often used in such circumstances as a means of estimating variability, but it should be noted that it depends on a series of independent variables which is only valid when there is one sample per animal. It is not strictly valid when more than one sample is taken from the same animal, so for those toxicokinetic protocols where more than one sample is taken, such as with ocular sampling, it would be more appropriate to employ mixed effect modelling.

Simple summary statistical representation of the data is usually acceptable unless some special comparisons are needed, but to do this with sufficient power to the prove the null hypothesis, a larger number of animals than normal are required. Perhaps the only statistical evaluation of toxicokinetic parameters required is to assess linearity of the toxicokinetics at different doses. These more advanced methods for sampling and data analysis, although not in present guidance, could form the basis of future amendments as done for MIST and validation guidances.

7.12 Report Writing

This aspect was not included in the original guidance but addresses questions which are often raised when reporting toxicokinetic results. When reporting bioanalytical and toxicokinetic data, it is much easier, for both the author and the reviewer, to have each included as a separate contributor report usually as an appendix to the main toxicology report even though the data may be generated by different laboratories. The main toxicology report should have a short summary of the main toxicokinetic findings, referring back to the full toxicokinetic report in the appendix.

The toxicokinetic report should reference full species matrix validation, how the analysis was performed, together with any statistical analysis employed (e.g. dose linearity) and provide any deviations from the protocol. The data including the individual concentration–time data and the toxicokinetic parameters from individual animals should be provided, together with descriptive statistics (e.g. mean ± SD, median and range), making full use of tables and figures. It is also useful for the reader to have a summary table of the exposure (C_{max} and AUC_{0-24h}) for the final dose, when steady state should have been reached, at the NOAEL. It should be remembered that a reader or reviewer of the report will want to find the data quickly and easily, so all the data should be rationally ordered and simply presented. The individual data are required to allow the reviewer to determine the variability, looking for outliers to see, for instance, if there is relationship to toxicological findings. The reviewer may even want to reanalyze the data, either using the methods employed in the report or using an alternative method for comparative purposes. Thus, all the data should be in an easily assessable form that will allow the reviewer to perform any additional evaluations (Table 7.3).

Interpretation in the toxicokinetic report is best kept to a minimum and only to those aspects contained within the report (e.g. linearity of pharmacokinetics or degree of accumulation). However, when the toxicokinetics is also being used to define the multiple-dose pharmacokinetics of a compound in the toxicological species, more data than the standard C_{max}, T_{max} and AUC for each dose are needed and should be included in the report such as accumulation ratio, peak to trough ratios, terminal half-life after the last dose and dose linearity of each of the doses measured, for both drug and relevant metabolites. Again, however, care should be taken not to over interpret the data. It should be remembered that the most important conclusions, safety margins, are contingent on data from outside the report including clinical studies, and these will only be meaningful when the final clinical dose(s) is known. Moreover, this may change as the drug development process proceeds, so it is best left to subsequent authors of the various regulatory summary documents, such as for the IND and NDA in the USA or IMPD and MAA in Europe and the Investigator's Brochure.

The preparation of these regulatory summary documents, however, requires some careful thought since it is bringing together data from many different sources. The safety margins are typically calculated from (7.3), and the exposure period is generally 24 h but may be longer for some very long half-life compounds (e.g. monoclonal antibodies), although the calculation is essentially the same:

$$\text{Safety margin} = \frac{\text{Steady}-\text{state } C_{max} \text{ or } AUC_{0\text{-}24h} \text{ at the NOAEL}}{\text{Steady}-\text{state clinical } C_{max} \text{ or } AUC_{0-24h}} \tag{7.3}$$

For small molecules, another major consideration in safety margin calculations is whether to use total drug concentrations or free (unbound) drug concentrations. Theoretically, it is only the unbound drug concentrations that are available to elicit a toxic response, and, as stated in the original ICH guidelines, the exposure measurements should, theoretically, be based on the unbound concentrations. However,

Table 7.3 Suggested contents, although not limited to, for inclusion in bioanalytical reports of toxicokinetic data and toxicokinetic reports

Bioanalytical reports	Toxicokinetic reports
Principal investigators' statement and signature	Principal investigators' statement and signature
QA statement where applicable	QA statement where applicable
Key personnel involved with the study	Key personnel involved with the study
Contents	Contents
Abstract	Abstract
Abbreviations	Abbreviations
Methods—Reference standards and matrix preparation – Study sample preparation and analysis – Analytical conditions – Acceptance criteria	Methods—Overview of experimental design – Toxicokinetic analysis including calculation and statistical methods
Results—Overview of analysis – Study sample data – Calibration data – QC sample data – Repeat analyses – Incurred sample reanalysis (ISR) if applicable	Results—Toxicokinetic parameters for parent compound with descriptive statistics – Toxicokinetic parameters for metabolites with descriptive statistics, if applicable – Additional statistical analyses, if applicable
Discussion (keep brief and related to data contained in report)	Discussion (keep brief and related to data contained in report)
Archiving of data	Archiving of data
References	References
Tables—Summary of runs – Study sample data – Calibration curve parameters – QC sample data – Sample reanalysis data – ISR data if applicable	Tables—Individual toxicokinetic parameters of parent compound and metabolites where applicable with descriptive statistics for each dose and sampling day – Additional statistical analysis (e.g. linearity of kinetics), if applicable – Individual plasma concentrations with descriptive statistics (parent compound and metabolites, if applicable) for each dose and sampling day
Figures—Calibration curves – Representative chromatograms including blanks	Figures—Individual plasma profiles of parent compound and metabolites where applicable for each dose combining first, mid- and last dose where available – Average plasma profiles with variability of parent compound and metabolites where applicable combining first, mid- and last dose where available – Linearity of kinetics

it should be noted that plasma protein binding should be based on the plasma protein binding, embracing the concentration range seen in the respective toxicological and clinical studies, since saturation of plasma protein binding, especially at the high concentrations reached in toxicology studies, is reasonably common.

Failure to take this into account could result in significant errors of the safety margins, especially for compounds showing high plasma protein binding.

In some instances where the in vitro toxicity has been measured, such as specific receptor ($5HT_2B$ cardiac binding) or transporter binding (e.g. hERG), the ratio of the toxic ED50 or ED20 to the unbound highest therapeutic plasma level at steady state can be used as another assessment of safety margin. However, the interpretation of such data is confounded by protein binding and tissue uptake which cannot be easily unravelled, but this approach may provide additional information to allow safety dosing decisions to be taken.

The extent of the desired safety margin will depend somewhat of the therapeutic area for the drug and the risk benefit ratio and the type of toxicity. Mortality weighs more heavily, for example, than liver hypertrophy, but as a general rule of thumb, margins greater than 30–50-fold seems a little high ×10 is usually the minimum with <10 for life threatening disease are deemed acceptable. For those compounds treating life-threatening diseases and for which there are very few, if any, alternatives, the safety margin tends to be lower. For this latter category, a tenfold safety margin or less may be acceptable. In practice, for the FIH, starting dose is based on MRSD and not necessarily on toxicokinetics, whilst the maximum dose in early human studies is dependent on the worst case parameter at the lowest NOAEL, independent of whether it is Cmax or AUC, bound or unbound, but can be exceeded if the human tolerates it better than the animals used in safety studies when 'safety margins' may be less than 1.

7.13 Conclusion

The initial S3A guidance published 16 years ago went some way to integrate the current science, the practices of the pharmaceutical companies and the regulatory needs, producing for the first time a framework for assessing drug exposure in animals rather than doses to compare with human. This resulted in agreements across the three regions on the type and extent of work required and produced simple workable guidance representing a major step forward at the time of writing. The large number of animal pharmacokinetic studies undertaken previously could be reduced, emphasising the need to mainly focus on toxicokinetics in pivotal safety studies, but the standard of analysis needed to be high with validated GLP methods. Guidance was given of what and when to measure analytes and which TK parameters should be described and how these could be used in the design of safety studies. The issue of the timings and the need for radioactive studies in man and tissue distribution in animals was not easily resolved due to the understandable concerns in Japan on the use of radioactivity in humans, so a subsequent guidance S3B was produced, but the emphasis on doing these studies early prior to phase 1, unless findings warrant it, is now questionable. Although the original S3A guidance was a useful starting point, it has been recognised that further clarification was necessary. The M3 guidance showed that toxicokinetics was probably of little use in defining the starting dose in man and we still needed scaled animal doses.

In principle, the top dose allowable in clinical studies should be based on safety margins calculated from the lowest calculated NOAEL, depending on the adverse effect, but no guidance has actually stated this or the circumstances when it would be possible to go higher. This still remains an issue in clinical development. Also there are the questions of what should be done with metabolites, when should they be measured, what should be measured and then what should be done with the results? The MIST guidance has gone some way to address these questions although there is some debate on how percentage 'exposure' should be calculated, but the relevance is still questionable since often it is reactive metabolites which are toxic and being short lived, the circulating levels are low (less than 10%) and urinary levels of potential end products are, in practice, difficult to interpret. In the intervening years, little progress has been made on how we should try and relate toxicokinetics to toxicodynamics which is clearly a key to understanding interspecies differences in the relationship between drug levels and toxicity and their extrapolating to man. Recent agreement on nephrotoxicity biomarker will hopefully pave the way for other acceptable indices of toxicity and perhaps further guidance will be forthcoming. Also it is hoped that there will be more translational research in non-invasive methods of assessing organ toxicity in real time and serve the dual purpose of linking exposure with toxicity and the amelioration of disease progression in clinical trials. Much discussions have taken place on bioanalysis and validation of methods to the nth degree, and yet even the 'simple' measurement of AUC warrants closer attention. Unfortunately the easy widespread use of kinetic programmes such as WinNonlin and SAS by scientists not necessarily versed in the detail of theoretical kinetics can lead to misuse and misinterpretation of the results. The improvements in the sensitivity of methods brought about by LS–MS/MS have allowed new sampling approaches to be investigated. The combination of the new developments in composite sampling, auto-sampling, microsampling with plasma separation, DBS analysis and population approaches to the data should go some way to achieve one of the key objectives of the ICH, in reducing the number of animals used in development by the use of serial concomitant sampling in rodents. However, additional validation of the DBS method in comparison with wet analysis needs to be undertaken before it achieves regulatory acceptance and is universally utilised throughout development. Perhaps dried plasma analysis would overcome many of the difficulties of not knowing the in vivo RBC/plasma ratios. Other issues of interpretation of data which were not addressed in the original guidance nor since may need to be considered in the future including the following: Should we routinely measure drug and metabolite levels in in vitro tests such as mutagenicity, cytotoxicity, hERG, Purkinje fibres or any other cell- or tissue-based safety assays? Should we routinely undertake protein binding at the described NOAELs and compare them with therapeutic levels in humans to define more accurately the top dose allowable in man, and if so, should they be ex vivo? What do we do with chiral compounds and their subsequent chiral metabolites? In the translation of species exposure in longer term safety studies, should we take into account the relative life span of the animals? Is a cumulative AUC exposure in, say, the rat over 6 months (25% of life span) the same as the cumulative AUC in

humans for the same period (4% of life span), and should adjustments be made to compensate for such differences? How much of what we do with small organic compound has any relevance to large proteins or monoclonal antibodies?

It is clear that despite the incompleteness of the S3A guidance, at the time, it set a much-needed conceptual basic framework to use and build on, and like the American constitution, interpretation of its basic tenets will continue as we learn more about how and when to use it. Already we have seen important improvements, and we should look forward to further advances particularly in the field of biologics. However, there has to be a careful balance between the benefit and risk in drug development, and can we run the risk of ever-increasing regulatory requirements to such an extent that good drugs are not developed and patients are denied the benefits of new medicines? Was there really a big problem before MIST clouded our eyes and potentially slowed our progress?

References

Aarons L, Graham G (2001) Methodological approaches to the population analysis of toxicity data. Toxicol Lett 120:405–410

Bailer AJ (1988) Testing for the equality of area under the curves when using destructive measurement techniques. J Pharmacokinet Biopharm 16:303–309

Baillie TA, Cayen MN, Fouda H et al (2002) Drug metabolites in safety testing. Toxicol Appl Pharmacol 182:188–196

Baranczewski P, Stancazak A, Sundberg K et al (2006) Introduction to *in vitro* estimation of metabolic stability and drug interactions of new chemical entities in drug discovery and development. Pharm Rep 58:453–472

Barr JR, Maggio VL, Patterson DG et al (1996) Isotope dilution – mass spectrometric quantification of specific proteins: model application with apolipoprotein A-I. Clin Chem 42:1676–1682

Barter ZE, Bayliss MK, Beaune PH, Boobis AR, Carlile DJ, Edwards RJ, Houston JB, Lake BG, Lipscomb BC, Pelkonen OR, Tucker GT, Rostami-Hodjegan A (2007) Scaling factors for the extrapolation of *in vivo* metabolic drug clearance from *in vitro* data: reaching a consensus on values of human microsomal protein and hepatocellularity per gram of liver. Curr Drug Metab 8:33–45

Batra VK (1995) Toxicokinetics/toxicodynamic correlations: goals, methods, and limitations. Toxicol Pathol 23(2):158–164

Beharry M. (2010) DBS: a UK (MHRA) regulatory perspective. Bioanalysis. Aug;2(8):1363–1364

Boxenbaum H (1982) Interspecies scaling, allometry, physiological time and the ground plan of pharmacokinetics. J Pharmacokinet Biopharm 10:201–227

Buchanan JB, Burke LT, Melnick L (1997) Purpose and guidelines for toxicokinetic studies within the national toxicology program. Environ Health Perspect 105:468–471

Buscher BA, Gerritsen H, van Scholl I, Cnubben NH, Brull LP (2007) Quantitative analysis of Tenecteplase in rat plasma using LC-MS/MS as an alternative for ELISA. J Chromatogr B Analyt Technol Biomed Life Sci 852:631–634

Campbell DB (1990) The development of chiral drugs. Acta Pharm Nord 2(3):217–226

Campbell DB (1995) The use of toxicokinetics for the safety assessment of drugs acting in the brain. Mol Neurobiol 11(1–3):193–216

Carrera G, Mitjavila S, Lacombe C, Derache R (1976) Toxicocinetique d'un pesticide du groupe des thioquinoxalines: L'oxythioquinox. Toxicology 6:161–171

CDER/FDA (2006) Guidance for industry bioanalytical method validation; Fit for purpose method development and validation for successful biomarker measurement

CDER–CBER–CDRH–FDA (2007) Guidance for industry: pharmacogenomics data submission

CHMP/EMEA (2008a) Biomarker qualification: guidance to applicants
CHMP/EMEA (2008b) Final report on the Pilot Joint EMEA/FDA/VXDS experience on qualification on nephrotoxicity biomarkers
Cody RB, Laramee RB (2005) Versatile new ion source for analysis of materials in open air under ambient conditions. Anal Chem 77:2297–2302
Cooks RG, Ouyang Z, Takats Z, Wiseman JM (2006) Ambient mass spectrometry. Science 311: 1566–1570
Crawford E, Gordon J, Wu J-T, Musselman B, Liu R, Yu S (2011) Direct analysis in real time coupled with dried spot sampling for bioanalysis in a drug-discovery setting. Bioanalysis 3:1217–1226
Dahlem AM, Allerheiligen SR, Vodicnk MJ (1995) Concomitant toxicokinetics: techniques for and interpretation of exposure obtained during the conduct of toxicology studies. Toxicol Pathol 23:170–178
Dedrick RL (1973) Animal scale-up. J Pharmacokinet Biopharm 1:435–461
Dong JQ, Salinger DH, Endres CJ, Gibbs JP, Hsu C-H, Stouch BJ, Hurh E, Gibbs MA (2011) Quantitative prediction of human pharmacokinetics for monoclonal antibodies. Clin Pharmacokinet 50:131–142
EMEA (2011) Guideline on bioanalytical method validation
Ezan E, Bitsch F (2009) Critical comparison of MS and immunoassays for the bioanalysis of therapeutic antibodies. Bioanalysis 1:1375–1388
Fast DM, Kelley M, Viswanathan CT, O'Shaughnessy J, King SP, Chaudhary A, Weiner R, DeStefano AJ, Tang D (2008) Workshop report and follow-up—AAPS workshop on current topics in GLP bioanalysis: assay reproducibility for incurred samples—implications of crystal city recommendations. AAPS J 11:238–241
FDA (2001) FDA guidance for industry: bioanalytical method validation. FDA, Washington, DC
FDA (2005) FDA guidance for industry: estimating the maximum safe starting dose in initial trials for therapeutics in adult healthy volunteers. FDA, Washington, DC
FDA (2008) FDA guidance for industry: safety testing of drug metabolites. FDA, Washington, DC
FDA (2009) FDA Guidance for Industry: M3(R2) Nonclinical safety studies for the conduct of human clinical trials and marketing authorization for pharmaceuticals, 2010; ICH Topic M3 (R2) Non-clinical safety studies for the conduct of human clinical trials and marketing authorization for pharmaceuticals. FDA, Washington, DC
FDA (2010a) FDA guidance for industry: nonclinical safety studies for the conduct of human clinical trials and marketing authorization for pharmaceuticals. FDA, Washington, DC
FDA (2010b) FDA guidance for industry: S9 Nonclinical evaluation for anticancer pharmaceuticals. FDA, Washington, DC
FDA Guidance (2011) Reproductive and developmental toxicities – integrated study results to access concerns
Garner RC, Lappin G (2006) The phase 0 microdosing concept. Br J Clin Pharmacol 61:367–370
Geber SA, Rush J, Stemman O, Kirschner MW, Gyqi SP (2003) Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. Proc Natl Acad Sci USA 100:6940–6945
Griffini P, James A, Roberts AD, Pellegatti M (2010) Metabolites in safety testing: issues and approaches to the safety evaluation of human metabolites in a drug that is extensively metabolized. J Drug Metab Toxicol 1:102
Grouzmann E, Cavadas C, Grand D, Moratel M, Aubert JF, Brunner HR, Mazzolai L (2003) Blood sampling methodology is crucial for precise measurement of plasma catecholamines concentrations in mice. Pflugers Arch 447(2):254–258
Hashizume T, Yoshitomi S, Asahi S, Uematsu R, Matsumura S, Chatani F, Oda H (2010) Advantages of human hepatocyte-derived transformants expressing a series of human cytochrome p450 isoforms for genotoxicity examination. Toxicol Sci 116(2):488–497
Hing JP, Woolfrey SG, Greenslade D, Wright PMC (2002) Distinguishing animal subsets in toxicokinetic studies: comparison of non-linear mixed effects modeling with non-compartmental methods. J Appl Toxicol 22:437–443
Hoshino-Yoshino A, Kato M, Nakano K, Ishigai M, Kudo T, Ito K (2011) Bridging from preclinical to clinical studies for tyrosine kinase inhibitors based on pharmacokinetics/pharmacodynamics and toxicokinetics/toxicodynamics. Drug Metab Pharmacokinet 26(6):612–620

ICH (1995a) ICH Topic S1C (R2): Dose selection for carcinogenicity studies of pharmaceuticals CHMP/ICH/383/95

ICH (2009) ICH Topic M3 (R2): Non-clinical safety studies for the conduct of human clinical trials and marketing authorization for pharmaceuticals

ICH (1994) Pharmacokinetics: guidance for repeated dose tissue distribution studies S3B

ICH (1995b) Note for guidance on toxicokinetics: the assessment of systemic exposure in toxicity studies S3A

Ings R (1989) Pharmacokinetics and its application to drug development. In: Illing HPA (ed) Xenobiotic metabolism and disposition: the design of studies on novel compounds. CRC, Boca Raton, FL

Ings R (1990) Interspecies scaling and comparisons in drug development and toxicokinetics. Xenobiotica 20:1201–1231

Ings R (2009) Microdosing: a valuable tool for accelerating drug development and the role of bioanalytical methods in meeting the challenge. Bioanalysis 1:1293–1305

Ingwersen SH, Kiehr B, Iversen L, Andersen MP, Petersen Y, Rytved KA (2002) Non-linear mixed effects modeling of sparse concentration data from rats: application to a glycogen phosphorylase inhibitor. Eur J Drug Metab Pharmacokinet 27:203–212

Ji QC, Rodila R, Gage EM, El-Shourbagy TA (2003) A strategy of plasma protein quantitation by selective reaction monitoring of an intact protein. Anal Chem 75:7008–7014

Kalvass CJ, Maurer TS, Pollack GM (2007) Use of plasma and brain unbound fractions to assess the extent of brain distribution of 34 drugs: comparison of unbound concentration ratios to *in vivo* P-glycoprotein efflux ratios. Drug Metab Dispos 35:660–666

Khor SP, McCarthy K, Dupont M, Murray K, Timony G (2000) Pharmacokinetics, pharmacodynamics, allometry and dose selection of rPSGL-Ig for phase I trial. J Pharmacol Exp Ther 293: 618–624

Kippen AD, Cerini F, Vadas L, Stocklin R, Vu L, Offord RE, Rose K (1997) Development of an isotope dilution assay for precise determination of insulin, C-peptide and proinsulin levels in non-diabetic and type II diabetic individuals with comparison to immunoassay. J Biol Chem 272:12513–12522

Kurawattimath V, Pocha K, Thanga Mariappan T, Trivedi RK, Mandlekar S (2012) A modified serial blood sampling technique and utility of dried-blood spot technique in estimation of blood concentration: application in mouse pharmacokinetics. Eur J Drug Metab Pharmacokinet 37(1):23–30

Lavé T, Chapman K, Goldsmith P, Rowland M (2009) Human clearance prediction: shifting the paradigm. Expert Opin Drug Metab Toxicol 5(9):1039–1048

Li F, Zulkoski J, Fast D, Michael S (2011) Perforated dried blood spots: a novel format for accurate microsampling. Bioanalysis 3(20):2321–2333

Mahmood I (2007) Application of allometric principles for the prediction of pharmacokinetics in human and veterinary drug development. Adv Drug Deliv Rev 59:1177–1192

Miida H, Arakawa S, Shibaya Y, Honda K, Kiyosawa N, Watanabe K, Manabe S, Takasaki W, Ueno K (2008) Toxicokinetic and toxicodynamic analysis of clofibrate based on free drug concentrations in nagase analbuminemia rats (NAR). J Toxicol Sci 33(3):349–361

Muller PY, Dieterle F (2009) Tissue specific, non invasive toxicity biomarkers; translation from preclinical safety assessment to clinical safety monitoring. Expert Opin Drug Metab Toxicol 5(9):1023–1036

Obach RS, Baxter JG, Liston TE, Silber BM, Jones BC, MacIntyre F, Rance DJ et al (1997) The prediction of human pharmacokinetic parameters from preclinical and *in vitro* metabolism data. J Pharmacol Exp Ther 283:46–58

Peck CC, Barr WH, Benet LZ et al (1994) Opportunities for integration of pharmacokinetics, pharmacodynamics, and toxicokinetics in rational drug development. J Clin Pharmacol 34(2):111–119

Peterson RA, Gabrielson KL, Allan Johnson G, Pomper MG, Coatney RW, Winkelmann CT (2011) Continuing education course #1: non-invasive imaging as a problem-solving tool and translational biomarker strategy in toxicologic pathology. Toxicol Pathol 39(1):267–272

Rowland M, Benet LZ, Graham GG (1973) Clearance concepts in pharmacokinetics. J Pharmacokinet Biopharm 1:123–136

Rueff J, Chiapella C, Chipman JK et al (1996) Development and validation of alternative metabolic systems for mutagenicity testing in short-term assays. Mutat Res 353(1–2):151–176
Shah VP (2007) The history of bioanalytical method validation and regulation: evolution of a guidance document on bioanalytical methods validation. AAPS J 9:E43–E46
Shah VP, Midha KK, Dighe SV et al (1992) Analytical methods validation: bioavailability, bioequivalence and pharmacokinetic studies. Pharm Res 9:588–592
Shah VP, Midha KK, Findlay JW et al (2000) Bioanalytical method validation – a revisit with a decade of progress. Pharm Res 17:1551–1557
Shin BS, Kim DH, Cho CY et al (2003) Pharmacokinetic scaling of SJ-8029, a novel anticancer agent possessing microtubule and topoisomerase inhibiting activities, by species-invariant time methods. Biopharm Drug Dispos 24:191–197
Smith C, Skyes A, Robinson S, Thomas E (2011) Evaluation of blood microsampling techniques and sampling sites for the analysis of drugs by HPLC-MS. Bioanalysis 3(2):145–156
Vogel JS, Turtletaub KW, Finkel R, Nelson DE (1995) Accelerator mass spectrometry. Anal Chem 67:A353–A359
Wong P, Pham R, Bruener B, James C (2010) Increasing efficiency for dried blood spot analysis: prospects for automation and simplified sample analysis. Bioanalysis 2:1787–1789
Wsol V, Skalova B, Szotakova B (2004) Chiral inversion of drugs: coincidence or principle. Curr Drug Metab 5:517–533
Zhong WZ, Williams MG, Branstetter DG (2000) Toxicokinetics in drug development: an overview of toxicokinetic application in the development of PNU-101017, an anxiolytic drug candidate. Curr Drug Metab 1(3):243–254

Chapter 8
Duration of Acute and Chronic Toxicity Testing in Animals (ICH S4A and S4B)

Per Spindler and Herman Van Cauteren

Abstract To support approval of pharmaceuticals for long-term use in humans, it is required that product safety is supported by acute and chronic toxicity studies in rodents and non-rodents. The duration of acute toxicity studies (S4A) and chronic rodent studies (S4B) was harmonised between the three ICH regions in 1991, whereas the process of harmonising the duration of non-rodent studies was initiated. The US FDA originally required studies of at least 12-month duration, whereas in Japan and EU studies of 6-month duration were considered acceptable as an ICH objective. In this chapter, the background for the ICH S4B guideline regarding the duration of non-rodent repeated-dose toxicity studies is explained and lessons learned are discussed. Since the guideline was issued in 1997, changes occurred in, e.g. the language of the European legislation, and the requirements for non-clinical studies to support clinical development have progressed within the ICH (M3): we therefore consider options such as prospective evaluation, biomarker-based mechanistic understanding, toxicokinetics and the use of evidence-based medicine to support further joint activities to harmonise the duration of non-rodent toxicity studies at the global level.

Abbreviations

CMR Centre for Medicines Research (now Centre for Innovation in Regulatory Science CIRS)
CRO Contract Research Organisation

P. Spindler (✉)
Biopeople, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
e-mail: pesp@biopeople.ku.dk

H. Van Cauteren
Pharmaparacelsus LLC, Gierle, Belgium

J.W. van der Laan and J.J. DeGeorge (eds.), *Global Approach in Safety Testing*, AAPS Advances in the Pharmaceutical Sciences Series 5, DOI 10.1007/978-1-4614-5950-7_8,

ECVAM	European Centre for the Evaluation of Alternative Methods
EFPIA	European Federation of Pharmaceutical Industries and Associations
EMEA	European Agency for the Evaluation of Medicinal Products (now European Medicines Agency EMA)
EU	European Union
EWG	Expert Working Group
FDA	United States Food and Drug Administration
ICH	International Conference of Harmonization of Pharmaceutical for Human Use
JPMA	Japanese Pharmaceutical Manufacturers Association
MHW	Japanese Ministry of Health and Welfare (now Ministry of Health Labour and Welfare, MHLW)
NIH	National Institutes of Health
PhRMA	Pharmaceutical Research and Manufacturers of America

8.1 Background

The First International Conference on Harmonisation in Brussels, 1991, had the perspective to identify areas where harmonization of testing requirements in Japan, the USA and EU could lead to reduced resources in the development of new pharmaceuticals. One of the major areas was represented by the testing for acute and long-term toxicity in rodents and particularly in non-rodents. FDA published the guideline in 1996 (Food and Drug Administration 1996). The topic of acute toxicity testing (S4A) was a "quick win" as pilot studies such as dose-range finding studies could be used to replace the need of full-size GLP single-dose toxicity studies, and it was agreed that the requirements for an LD_{50} determination in rodents or non-rodents could be replaced by well-designed tolerance studies in either two rodent species or one rodent and one non-rodent species (Hayyashi 1993; Casciano 1995). These changes to the recommended testing battery to support safety before first human exposure was further supported by a documented impressive safety record of phase I clinical trials (Spindler et al. 2000).

Before the start of processes of ICH, in the mid-1980s, the duration of chronic toxicity studies was not harmonised between regions and countries. Depending on the country to which a marketing authorisation application of a new pharmaceutical was submitted, the requested duration of chronic testing could vary between 6, 12 or even 18 months. The historical timelines for the establishment of the ICH S4 guideline are presented in Table 8.1.

In 1988, the JPMA performed a questionnaire survey of chronic toxicity studies (Igarashi 1993). The questionnaire asked the member companies what toxicities were newly found in repeated-dose toxicity studies with 12-month treatment. Fifty (50) companies replied to the questionnaire referring a total of 124 long-term toxicity studies (12 months and longer) in rats, dogs or monkeys, which were completed studies since 1983. Other surveys were carried out by the FDA and the Centre for Medicines Research (CMR).

The FDA assessed approximately 30 cases for which both 6- and 12-month non-rodent toxicity studies were available for comparison. In some cases, the FDA stated

Table 8.1 Historical timelines for the establishment of the ICH guideline for the duration of acute and chronic toxicity testing (S4A and S4B)

1991	ICH 1, Brussels, EU • Agreement on single-dose toxicity testing and abolishment of the LD50 determination (ICH S4A) • Harmonization of the duration of the chronic rodent study to 6 months from either 6 and/or 12 months • No agreement was reached and further data analysis was discussed (S4B)
1993	ICH 2, Orlando, USA
1995	ICH3, Yokohama, Japan
1996	ICH S4B Step 1 guideline released
1996	S4B EWG meeting
1997	S4B EWG Regulatory Assessment Group, Copenhagen • Recommended 9-month duration from data evaluations
1997	ICH S4B Step 2 guideline released
1997	ICH 4, Brussels, EU
1998	ICH S4B Step 4 guideline released
1999	ICH S4B Step 5—regional implementation of the guideline

that animal findings after 6 months were significantly enough to be included in labelling and resulted in a shift in the dose–response for a serious toxicity that could alter the estimated clinical margin of safety for the pharmaceutical. Differences were also detected between 6- and 12-month studies as during the IND process, clinical trials were altered and/or temporary suspended (DeGeorge et al. 1999).

The CMR conducted a voluntary questionnaire among pharmaceutical companies with a request to provide an assessment of 6- vs. 12-month non-rodent toxicity studies (DeGeorge et al. 1999; Parkinson 1992). Out of more than 100 cases of 6- and 12-month toxicity studies available, two cases with new noteworthy toxicological findings (with consequences of development of the pharmaceutical) at 12 months were identified.

There were no regional agreements on the requirements for the duration of non-rodent toxicity studies at this explorative stage of the ICH process. Japan and EU were satisfied with 6-month duration. The FDA argued that studies with a duration of 12 months might be needed based on the (5 or) 6 published cases. The EFPIA pharma experts recognised some of the cases, and some experts proposed that a prospective study on the six cases be conducted. However, such a prospective study would have a high cost of 2 MUSD or more, and furthermore, there was no consensus within the industry. After ICH 1 (1991), the FDA published their cases (Contrera et al. 1993).

At ICH 1 the participants agreed on a single duration for the chronic toxicity study in rodents of 6 months. In part this agreement was achieved by noting that this 6-month period of treatment covered a significant portion of the rodent's life span and that carcinogenicity studies of 2-year duration were also conducted in rodent as it was the experience and opinion that no noteworthy significant toxicity findings were detected in rodent in 1-year chronic toxicity studies that were not detected at 6 months or as part of the 2-year carcinogenicity study bioassay.

In preparation of the ICH 1 conference in Brussels among other working groups, a S4B Expert Working Group (S4 EWG) for safety assessment was established. The first meeting of the S4B EWG was held in 1992 and was a critical review of information

from FDA, CMR and JPMA databases; discussions were quickly halted since the FDA indicated that on the basis of their interpretation of the data, 12-month dog studies would continue to be required. Attempts were made to design a prospective study to assess whether 12 and also 9 months were needed as compared to 6-month studies. This attempt was, however, unsuccessful due to lack of consensus of the value of such a study and the resources involved in running it.

At ICH 3 in Yokohama, December 1995, it was decided by the ICH Steering Committee to continue the S4B topic and prepare the topic for presentation in 1997 at ICH 4 (Strandberg 1995). The regions maintained their original positions, and in order to move the topic forward, it was agreed that it would be valuable to address the issue again since additional data from adequately conducted 6- and 12-month toxicity studies had become available for analysis. An alternative approach was, however, also discussed, namely, as indicated above, to determine whether an altered study duration would address the findings that were of concern to the FDA and might be acceptable to all parties.

As already indicated, five special cases/substances from the FDA database were published in the early 1990s (Contrera et al. 1993). Following discussions about conducting new prospective studies, the FDA reassessed their database and released a discussion paper with 16 cases. The FDA provided a scientific background to propose duration of 9 months vs. 6 and/or 12 months. In their analysis they determined that there existed sufficient data from *in-life* and early decedent observations to conclude that while 6 months of treatment was not adequate, 12 months was not necessary.

The next special Expert Working Group meeting with experts from the six ICH partners (FDA, MHW, EU, EFPIA, JPMA, PhRMA) took place in Vienna in 1996. The method of reassessment of the 16 cases was discussed, and willingness of industry to provide the cases was expressed (Box 8.1). Based upon the review of the FDA database in preparation of ICH 1 and some additional cases, it was decided in 1996 to assess 16 cases with potentially different findings between 6 and 12 months of dosing in non-rodents, mainly dog studies (Table 8.2). The weight of evidence was to be re-analysed for their potential impact on the safety profile of the involved pharmaceutical. The reassessment was performed by the regulatory experts from the three authorities of Japan (MHW), the USA (FDA) and EU and coordinated by the European Agency for the Evaluation of Medicinal Products (EMEA). It was considered of pivotal importance that confidentiality was respected and it was agreed that only coded names for the drugs were used and their pharmacological class indicated: neither the name of the pharmaceutical nor the company would be disclosed. Thus, it was ensured that the re-evaluation would not have any potential impact on the development or marketing status of the particular pharmaceutical (Box 8.1). The regulators initiated reassessment of the 16 cases and met again in Copenhagen in January 1997 to finally review the cases and draft the step 2 document for ICH 4. It was noted that the US requirements for 12-month repeated-dose toxicology testing were related to clinical trial requirements and authorisation, while in Japan and EU, 6-month repeated-dose toxicology testing is used at the marketing authorisation, although it also may be requested for long-term clinical trials. Options for 9-month harmonization would be explored. The potential advantages and disadvantages of a harmonised maximum duration of non-rodent toxicity studies are presented in Table 8.3.

Table 8.2 Pharmacological classes of the 16 cases assessed

• Anxiolytic (2)
• Antidepressant
• Antidementia
• Antipsychotic
• NSAID (2)
• Beta-blocker
• Calcium channel blocker (3)
• Antihypertensive other
• Anti-androgen
• Anticoagulant
• Anti-osteoporosis
• Recombinant protein

Box 8.1 Assessment Criteria of the 16 Cases 6- Versus 9- Versus 12-Month Repeated Dose Toxicity Studies

Requested material from companies for review of the 16 cases:

(a) A copy of the full "repeated-dose toxicology non-rodent" report, with or without individual data for 6- and 12-month studies and 1- and/or 3-month studies on dogs (or other non-rodents)
(b) A short product profile of the pharmaceutical product including information on indication and clinical safety profile
(c) An opinion of the company (expert report or similar), e.g. from the registration documentation on the dog findings with respect to human safety

The EWG regulatory experts agreed to assess the 16 cases with a scientific approach and include the following information in reporting to the group:

1A. For what indication was/is the compound used/intended for use?
1B. Which dosage pattern (e.g. daily for lifetime, daily for months/years) would be relevant or intermittent (for how long time/with how long intervals to be foreseen).
2A. Main observations of toxic effects and their time of occurrence after start of dosing (weeks/months, how many cases occurring at which rate). Particularly the observation after 6, 9 and 12 months of dosing should be indicated, and 1- and/or 3-month data could be included.
2B. Please observe that although the main finding of an effect occurs after 12 months, there may be low frequency (minor changes) to be observed already after 6 months of dosing or between 6 and 12 months. Then describe the occurrence pattern.
3A. Are there any new findings at 12 months.
3B. Are there shifts in the dose–response curve for some effects when comparing 6- and 12-month data.
3C. Are these differences leading to a change in the risk assessment based on these studies.
4. Are there indications of reversibility of changes observed.

Table 8.3 Advantages and disadvantages a harmonised maximum duration of non-rodent toxicity studies

Scenario	Feasibility	Advantage	Disadvantage
9 months in Japan, the USA and EU	Not easy for Japan and EU (because 6 months is acceptable) Further scientific basis may be insufficiently known Not accepted in the USA	Animal welfare—estimated 25–30 % savings One study instead of two No international discrepancies Support timely progression of multiregional clinical trial programmes	Impression of being only compromised if not supported by adequate scientific data Requires the generation of a new historical control database
9 months in Japan and EU 12 months in the USA	Easy If Japan and EU block attempts to compromise, the FDA likely withdraw its proposal for harmonization	Scientific evidence from most retrospective databases prepared for ICH1 is dominating Maintains the (non-harmonised) status quo	Animal welfare Two studies for same objective Potential for different safety margins between regions and thus differences in product labelling Limited scientific progress Only duplicating repeated-dose studies may generate new discrepancies

The tripartite regulators came to a mutual understanding at the meeting in Copenhagen, January 1997. The difference in the length of repeated-dose toxicity studies in non-rodent animals represented a major problem in the harmonization endeavours of ICH, and the topic is closely connected to the timing of non-clinical safety studies in relation to clinical trials. The inability to harmonise in these areas runs counter to one of the main ICH objectives to reduce the amount and length of experimentation on animals. At the closing plenary session of ICH 3 in 1995, it was concluded that a way forward was to get access to the data for all 16 problem drugs identified to date, regardless of file (EFPIA, JPMA, FDA, CMR), to have that data analysed be the regulatory authorities in the three regions and then to hold a tripartite meeting to come up with a recommendation for a harmonised position on the duration of repeated-dose toxicity studies in non-rodents (Strandberg 1995). There was no complete agreement among the regulators with respect to the comparability in design and conduct of 6- and 12-month studies for some of the cases to allow assessment of the differences in study results. There was also varying degrees of concern for the differences in findings detected between the studies of different durations. However, it was considered that there was sufficient information that a

9-month study duration would be adequate to capture the long-term toxicity findings identified in the 16 cases examined. Therefore, in consideration of the data reviewed and recognising the need to (1) speed the development of effective and safe pharmaceuticals, (2) eliminate waste of animal and financial resources through conduct of partially duplicative studies and (3) have an internationally recognised standard that adequately protects the public as perceived within each regulatory region, a 9-month non-rodent study duration was proposed by the regulatory members of the EWG S4B (Photo 8.1). During the discussions it was stated that while duration of 12-month testing was not necessary, 6-month studies could be insufficient to detect all toxic phenomena. Thus, there was no agreement on the clinical relevance of the findings. The 9-month maximum duration would facilitate global development and availability of new medicines while maintaining safeguards on regulatory obligations to protect human health. Avoiding unnecessary duplication of studies between regions, it was expected that the number of non-rodent animals was reduced by approximately one-third, e.g. it was expected that two repeated-dose toxicity studies with interim assessment would replace three studies in the non-clinical development of new pharmaceutical candidates (Van Cauteren and Lumley 1997).

In EU, studies of 6-month duration in non-rodents were acceptable according to Council Directive 75/318/EEC, as amended; and therefore, in consideration of the legislative power of EU directives versus "recommendatory" guidelines, the CPMP was ultimately not in agreement to increase the duration of studies from 6 to 9 months in EU, and in addition, also Japan had the position that 6-month studies were sufficient. Despite this lack of justification from a European and Japanese viewpoint, there were considerations—at global level—to avoid duplication of studies, and therefore the move from 12- to 9- month duration by the US could be considered as a slight improvement and a step towards harmonization. The three ICH industry partners (JPMA, PhRMA, EFPIA) and the authorities in Japan (MHW) and EU therefore supported a harmonised approach of the 9-month duration. Outside the USA, it was noted that where studies with a longer duration have been conducted, it would not be necessary to conduct a study of 6 months. It was proposed to ICH 4 in 1997 that a maximum of 9-month studies in non-rodents should be endorsed by all ICH parties (Sjöberg 1997; McClain 1997) (Photo 8.2).

The guidance specifically excluded biotechnology-derived pharmaceuticals, which were to be covered by ICH topic S6. The ICH M3 guideline that was under parallel discussions and cross-referenced the duration of *chronic* toxicity studies with ICH S4B.

The draft ICH S4B guideline was released for public consultation with the particular remark from the FDA that in some cases there was no complete agreement among the regulators with respect to the comparability in study design and conduct to allow assessment of whether there were differences in the findings at 6 and 12 months due to duration of treatment alone (Food and Drug Administration 1997). It was specified in the step 2 ICH consultation texts that studies of 12-month duration were *usually* not necessary and studies of shorter duration than 9-month duration may be sufficient and could be supported scientifically.

Photo 8.1 The ICH S4B EWR Regulatory Group agrees to a harmonised 9-month position for the duration of repeated dose non-rodent toxicity studies, Copenhagen, 14 January 1997 [*From left to right*: Drs. Mishihito Tachahashi (MHW), Laraine L. Meyers (FDA), Klaus Olejniczak (EU), Per Sjöberg (EU), Per Spindler (EU) and Joseph DeGeorge (FDA)]

Photo 8.2 International cooperation! A happy group of regulatory and industry experts celebrating what they thought was to be the final ICH S4A text on duration of chronic toxicity testing in animals (rodent and non-rodent toxicity testing), Narita, Japan, March 1997. Members of the Expert Working Group were Jens S. Schou (EU, Rapporteur), H. van Cauteren (EFPIA), Mishihito Takahashi (MWH), T. Igarashi (JPMA), Joseph DeGeorge (FDA) and Gregory S. Probst (PhRMA) (Reproduced with permission from Letter from the Editor (Jens S. Schou) Pharmacol Toxicol (1999) 84:1–2)

As discussions progressed in 1998, the FDA specified its position recognising that 9-month non-rodent toxicity studies are acceptable for marketing authorisation of most pharmaceuticals; however, specifying situations for which 6- and 12-month toxicity

studies may be appropriate. Six-month studies may be especially suited for clinical indications with short-term intermittent exposure, which might otherwise qualify for chronic testing, e.g. most bacterial infections, migraine, erectile dysfunction and herpes, and for indications of life-threatening disease with substantial long-term human clinical trial safety data available, e.g. cancer chemotherapy in advanced disease or in adjuvant use.

The FDA considered that 12-month studies may be appropriate for chronically used pharmaceuticals to be approved on the basis of short-term clinical data employing surrogate markers, e.g. AIDS therapies, where long-term clinical trial experience is limited. Twelve-month studies may also be useful for new pharmacological classes with limited experience for long-term toxic potential for the class, e.g. the first pharmaceutical in a new class. FDA also considered that pharmaceuticals used chronically and continuously for a substantial portion of a lifetime where a significant fraction of the patient population includes paediatric patients with normal life expectancy, e.g. treatment of epilepsy, sickle cell anaemia and asthma, also may be more appropriate evaluated in a 12-month study when there is no substantial prior market experience. These remarks were considered to be applied in the USA only.

ICH step 5 of the guideline implementation was reached in September 1998. In EU the guideline was adopted by CPMP in November 1998 (CPMP/ICH/300/95) (European Medicines 1999a; CPMP 2000; ICH 1998), by MHLW in April 1999 (lyaku-sin No. 655), and it was published by FDA in the Federal Register (FDA 1999). The 9-month duration of repeated-dose toxicity studies was considered acceptable in the final ICH guideline of Japan and EU (European Medicines 1999b). To avoid potential duplication of studies, wordings were included to explain that in cases where studies with longer duration than 6 months have been conducted, it would not be necessary to conduct a study of 6 months to comply with Council Directive 75/318/EEC in EU. In practice, FDA's considerations for the need of studies longer than 9 months, in particular for pharmaceuticals in new pharmacological classes (Food and Drug Administration 1999), set back the proposed harmonised maximum acceptable duration of 9 months between the three regions. Thus, for the USA, the power of decision for the duration of the non-rodent toxicity studies was given back to individual consultation with the divisions of the FDA.

This activity was the last one under the S4 topic of ICH. This overview dealing with the 1990s has some important follow-up in the next decade of the new millennium mainly because of updating and revising the ICH M3 guidelines (M3R2 published in 1999); this information updating chronic toxicity testing will be integrated in the next section.

8.2 Lessons Learned

When looking back at ICH S4B harmonization activities during period from 1991 to 1999 and considering current scientific insights and practices, some lessons learned could be brought forwards.

8.2.1 Reposition of Cases

Retrospective analysis on the cases reviewed by FDA (Contrera et al. 1993) suggests that some cases can be repositioned. Out of the cases mentioned in this publication, we have insights in cases, of which each can have a different interpretation today.

8.2.1.1 Reposition of "Case 13"

The serotonin reuptake inhibitor for the treatment of depression referenced as case 13 in DeGeorge et al. (1999) caused mortality in dogs at high dose levels in a 12-month repeated-dose toxicity study, whereas no mortality was observed in an earlier conducted 6-month study. In a subsequent paper, the mortality in the longer study, which was conducted in another CRO, was proved to be related to QTc prolongation mainly due to a dog-predominant didemethyl metabolite. This was confirmed by re-evaluation of ECG's from the 12-month study (information available from the US Pharmacopoeia).

In patients, this metabolite is much less produced and the QTc interval is not prolonged. The discrepancy of no mortality in the 6-month and the mortality in the 12-month dog study are proposed to be due to more stressful conditions in the different facilities where the 12-month study was conducted.

From the current insights and practices, the endpoint of possible mortalities in dogs dosed with this pharmaceutical is seen as a "bridge too far" and not necessary. Safety pharmacology studies in vitro and in vivo (ICH S7A and S7B) are now in place to assess after a single dose the cardiovascular safety such as QTc prolongation. Consequently, a thorough QTc study was conducted in humans and detailed monitoring of QTc in patients given the pharmaceutical, and results were negative. It is the most usual approach that the single-dose QTc prolongation in dogs would be confirmed in repeated-dose conditions (1, 3 months or longer). However, from a regulatory and ethical point of view, it is not recommended to induce QTc-related mortalities in dogs. Importantly, in this context, the guidelines for QTc were developed after these S4B discussions, and noteworthy so, assessment of ECG changes requires repeated-dose studies to capture the required safety information.

Assessment of QTc prolongation as such does not require chronic toxicity studies; however, since 6- and 12-month studies were originally done with this pharmaceutical, it was important to re-evaluate this case in a tripartite regulatory working party in 1997, and in 1997, this case supported the conclusion that the mortalities in the 12-month study did occur before the end of 9 months.

8.2.1.2 Reposition of "Case 15"

Another of the cases, a nonsteroidal anti-inflammatory drug (NSAID) referenced as case 15 in DeGeorge et al. (1999). Two dog repeated-dose toxicity studies were conducted: one with 6-month duration and one with 12 months including 6-month interim sacrifice.

Mortalities were described in Contrera et al. (1993) and DeGeorge et al. (1999) to occur after 6 months of treatment. However, what has not been disclosed in these publications is the fact that three dogs died or were sacrificed during the third month of treatment due to gastric ulceration and one of the dogs showed interstitial nephritis at histopathology. In the fourth dog that was sacrificed in the seventh month of treatment, slight interstitial nephritis was noted as well. From a publication point of view, in the comprehensive FDA paper, these details were not provided, and therefore, unfortunately, the mortalities appeared to have occurred only in the chronic toxicity study of 12 months (and not in the other study of 6 months).

From a mechanistic point of view, these dogs suffered from fatal and nonfatal gastric ulceration leading to episodes of local and systemic infections affecting also the kidneys, and the slight nephrotoxicity observed in these studies was considered to be secondarily related to the NSAID-related gastrointestinal toxicity at high toxic dose levels.

Moreover, urinalysis was not included in the comprehensive FDA paper, but especially microscopic analysis of the urine showed slight increase of hyaline casts at months 3, 6 and 12. Today, predictive and more specific biomarkers would be included in periodic urinalysis as well to increase the sensitivity of the analysis of nephrotoxicity.

The kidney toxicity reported in patients, as mentioned in the FDA paper, is a rather acute condition, and it is most likely related to competitive inhibition of uric acid excretion in the kidneys (Vaughan and Tucker 1987; Rossi et al. 1988; Harter 1988). However, this mechanism is not present in Beagle dogs questioning the overall relevance of the above-mentioned nephrotoxicity in the 12-month dog study.

8.2.2 Prospective Evaluation

In general, the predictive power of retrospective analysis of cases is inferior to conducting a prospective (cohort) study where only one variable is allowed, i.e. the duration of administration. All other variables such as time frame, batch, formulation, laboratory, testing facility, dog origin and housing and caretakers are kept similar. The proposal of retesting the pivotal cases this way by an industry effort was not easy to realise, and finally it did not happen. It should be mentioned here that the opinions of the ICH partners were divided as pro and contra including within the industry organisations. New cases identified after 1997 as presented by, e.g. industry representatives (Van Cauteren et al. 2000) could be taken into consideration if the path of prospective evaluation would be chosen.

8.2.3 "Public Positioning"

The acknowledgement of evidence-based medicine, known as a decision-making process that uses the best available objectively assessed knowledge as a substitute for an authority-based opinion, has progressed in the past decade. In the scientific

assessment of the safety of pharmaceuticals, the regulatory authorities and industry parties active in pharmaceuticals can learn from more recent, other related activities in academia, such as "evidence-based toxicology", derived from "evidence-based medicine".

Toxicologists may fail when shortcomings in the evidence necessitate reliance upon authority-based opinions rather than evidence-based conclusions (Guzelian et al. 2005). This new thinking has been discussed by ECVAM and others such as during the International Forum Towards Evidence-Based Toxicology where attempts are made to, e.g. improve in vitro testing predictability (Guzelian et al. 2009). This approach could also be applied to in vivo toxicology testing as well in the future.

In this respect, from a tactical and psychological point of view, another lesson learned is that "public positioning" of the opinion of one party is not a best/ideal first step in ICH. The publication of the five cases by FDA (Contrera et al. 1993), irrespective of its scientific merit, potentially put the other five ICH parties in a more difficult position in the following efforts to reach a harmonised maximum duration of studies.

8.2.4 Animal Welfare and Ethical Aspects of Preclinical Toxicology Testing

Harmonization of international toxicity testing guidelines for pharmaceuticals might have contributed to refinement and reduction of animal use. For single-dose toxicity testing, the ICH S4 guideline started with a quick win in dropping the LD_{50} testing and allowing dose (pilot) escalation studies to be brought instead. For repeated-dose toxicity testing, the duration of chronic testing in rodents was harmonised to 6 months (and no longer 12 months) in the three regions. For non-rodents, chronic toxicity studies were possibly harmonised when 9-month duration would have been acceptable but for certain pharmaceuticals. However, FDA could still request 12 months of duration in their guidelines drafted in 1997–1999 (FDA 1999). This harmonization of the ICH S4 guideline had as before mentioned the potential reduce the number of animals by approximately one-third, which is a large achievement (Van Cauteren and Lumley 1997): in some cases, it would, e.g. be possible to conduct only 1- and 9-month non-rodent toxicity studies, which clearly would lead to a significant reduction in animal use (Broadhead et al. 2000).

8.3 Next Steps for the Current Century and Future Outlook

After the ICH S4 guideline developments 1997–1999, new drug development was subjected to and opinion of the FDA divisions, which led to a variety of chronic toxicities in non-rodent studies with 6 plus 12, 6 or 9 months of duration.

Since the new century has started, the human genome has been published, increasing the value of genomics and proteomics mechanism-based biomarkers are much more applied increasing the sensitivity of the toxicity studies in animals and the efficacy and safety monitoring of patients in clinical trials. Great progress has also been obtained in the understanding of the importance of target organ toxicokinetics at the cellular level, i.e. assessment toxicokinetics linearity, saturation phenomenon and accumulation of cellular toxicity over the duration of repeated-dose administration. In addition, a battery of safety pharmacology studies have been harmonised during ICH 5 in 2000 introducing more functional testing in a cross-species manner and more short-term toxicity studies, e.g. limited dose escalation studies.

In 2001, a new European Directive 2001/83/EC was released. This directive is also addressing the duration of repeated-dose toxicity testing (Box 8.2). The wording of this directive says, "normally the duration of non-rodent repeated-dose toxicity studies shall last 3–6 months", and it no longer excludes the option for a duration of 9-month studies. This newer directive is much more subtle and flexible as compared with the older Council Directive of 1975 (75/318/EEC), which mandatorily stated a maximum duration of non-rodent repeated-dose toxicity studies of 6 months.

When checking current practices in the pharmaceutical industry, the duration of 9 months is much more utilised, and this trend is supported by the revision of the ICH M3 guideline published as step 4 in 2009 (International Conference on Harmonization 2009). The revised ICH M3 guideline provides recommendations for non-clinical studies to be conducted to support conduct of clinical trials and non-clinical studies to support a Marketing Authorisation Application: a duration of 9 months of non-rodent repeated-dose toxicity studies are explicitly recommended be it with important footnotes also allowing studies of 6-month duration in the 3 regions.

Box 8.2 The Text on the Duration of Repeated Dose Toxicity Studies in the European Legislation

Directive 2001/83/EC of the European Parliament and of the Council of 6 November 2001on the community code relating to medicinal products for human use:

(*4.2.3.b*) *Repeat-dose toxicity*: *Repeated dose toxicity tests are intended to reveal any physiological and/or anatomo-pathological changes induced by repeated administration of the active substance or combination of active substances under examination*, *and to determine how these changes are related to dosage*. *Generally*, *it is desirable that two tests be performed*: *one short--term*, *lasting 2–4 weeks*, *the other long-term*. *The duration of the latter shall depend on the conditions of clinical use*. *Its purpose is to describe potential adverse effects to which attention should be paid in clinical studies*. *The duration is defined in the relevant guidelines published by the Agency*.

According to ICH M3 R2 (International Conference on Harmonization 2009), the recommended duration of toxicity studies is generally equivalent to the duration of the intended clinical study, but with a minimum of 2 weeks and a maximum of 6 months in rodents and, previous to the revised guideline, 9 or 12 months in non-rodents, depending on the region in which the research is being carried out. A key point of harmonization in ICH M3(R2) is the setting of a maximum duration of 9 months for non-rodent toxicity studies, which will support both clinical studies and marketing authorisation for chronic-use therapeutics. This change may promote the elimination of a non-rodent study, because it is now feasible to conduct an additional 6-month study or a 6-month *interim* report of a 9-month study rather than delay long-term clinical trials until completion of a 9- (or 12) month repeated-dose toxicity study.

Unpublished discussions during the revision of the ICH M3 guideline indicate that new cases of chronic toxicity studies in dogs and primates provide evidence of toxicities detected in prior to 7 months of administration but at less than 9 months; thus, evidence that at glance is supportive of a 9-month recommendation. However, provision of any new case or new evidence would need in-depth scientific and regulatory scrutiny before concluding on the relevancy of the case for the duration of non-rodent repeated-dose toxicity studies.

Basically, "natural" harmonization of chronic toxicity testing in the non-rodent species was allowed to take place during the past 10 years. Scientific evidence and historical control data of 9-month repeated-dose toxicity studies are building up, and mechanistic understanding of toxicity following long-term administration is increasing.

Obviously, a potential reopening of the ICH S4B topic would require resources from both industry and regulators. A starting point could be collection and evaluation of cases of 6-, 9- and 12-month studies conducted in the time span since 1997 and initiated by a questionnaire to regulatory authorities and industry, respectively. The evaluation part of the exercise could include, e.g. analysis of potentially duplicated studies and the toxicokinetics profiles of parent compound and metabolites. To organise such a revisit of ICH topic S4B, one could envisage the formation of larger partnerships between the involved stakeholders and involving at least all three ICH regions. Partnering and funding options could be considered in programmes such as the Innovative Medicines Initiative and/or the next framework programme (Horizon 2020) of Europe, future programmes of NIH and/or the National Academies in the USA.

Acknowledgement We wish to acknowledge our colleagues Prof. Emeritus Jens S. Schou, Dr. Steven Spanhaak and Dr. Per Sjöberg for review and valuable discussion of this chapter.

References

Broadhead CL, Betton G, Combes R, Damment S, Everett D, Garner C, Godsafe Z, Healing G, Heywood R, Jennings M, Lumley C, Oliver G, Smith D, Straughan D, Topham J, Wallis R, Wilson S, Buckley P (2000) Prospects for reducing and refining the use of dogs in the regulatory toxicity testing of pharmaceuticals. Hum Exp Toxicol 19:440

Casciano DA (1995) Progress since ICH 2. In: International conference on harmonization of technical requirements for registration of pharmaceuticals for human use (ICH 3 1995), Proceedings of the third international conference on harmonization, Yokohama, pp 205–211

Committee for Proprietary Medicinal Products (CPMP): Note for guidance on repeated dose toxicity CPMP/SWP/1042/99 corr., 27 July 2000

Contrera JF, Aub D, Barbehenn E, Belair E, Chen C, Evoniuk G, Mainigi K, Mielach F, Sancilio L (1993) A retrospective comparison of the results of 6 and 12 months non-rodent toxicity studies. Adverse Drug React Toxicol Rev 12(1):63–76

DeGeorge JJ, Meyers LL, Takahashi M, Contrera JF (1999) The duration of non-rodent toxicity studies for pharmaceuticals. Toxicol Sci 49:143–155

European Medicines Agency (1999a) ICH topic S 4 Duration of chronic toxicity testing in animals (rodent and non rodent toxicity testing) Step 5. Note for guidance on duration of chronic toxicity testing in animals (rodent and non rodent toxicity testing (CPMP/ICH/300/95))

European Medicines Agency (1999b) CPMP/SWP/1042/99 Rev 1 Corr. Guideline on repeated dose toxicity

Food and Drug Administration (1996) Single dose acute toxicity testing for pharmaceuticals; revised guidance; availability. Fed Reg 61(166):4392–4393

Food and Drug Administration (1997) International conference on harmonization: Draft guidance on the duration of chronic toxicity testing in animals (rodent and nonrodent toxicity testing); availability. Fed Reg 62(222):61513–61515 [Docket No. 97D-0444]

Food and Drug Administration (1999) International conference on harmonization; guidance on the duration of chronic toxicity testing in animals (rodent and nonrodent toxicity testing); availability. Fed Reg 64(122):34259–34260 [Docket No. 97D-0444]

Guzelian PS, Victoroff MS, Halmes NC, James RC, Guzelian CP (2005) Evidence-based toxicology: a comprehensive framework for causation. Hum Exp Toxicol 24:161–201

Guzelian PS, Victoroff MS, Halmes C, James RC (2009) Clear path: towards an evidence-based toxicology (EBT). Hum Exp Toxicol 28:71–79

Harter JG (1988) Acute flank pain and hematuria: lessons from adverse drug reaction reporting. J Clin Pharmacol 28:560–565

Hayyashi Y (1993) Report on progress since ICH 1: single-dose toxicity, repeated-dose toxicity and carcinogenicity. In: International conference on harmonization of technical requirements for registration of pharmaceuticals for human use (ICH 2 1993), Proceedings of the second international conference on harmonization, Orlando, pp 31–33

ICH harmonised tripartite guideline: duration of chronic toxicity testing in animals (rodent and non-rodent toxicity testing) S4, Current Step 4 Version, dated 2 Sept 1998

Igarashi T (1993) A review of the Japanese Pharmaceutical Manufactures' Association database currently established to examine retrospectively the value of long term animal toxicity studies. Adverse Drug React Toxicol Rev 12(1):35–52

International Conference on Harmonization (2009) ICH Guideline M3 (R2): Guidance on nonclinical safety studies for the conduct of human clinical trials and marketing authorization for pharmaceuticals

McClain M (1997) Panel discussion session 3: Timing of non-clinical studies. In: International conference on harmonization of technical requirements for registration of pharmaceuticals for human use (ICH 4 1997), Proceedings of the fourth international conference on harmonization, Brussels, pp 309–316

Parkinson C (1992) Twelve month non-rodent studies debated at international forum. CMR News 10(2):1–3

Rossi AC, Bosco L, Falch GA, Tanner A, Temple RT (1988) The importance of adverse reaction reporting by physicians. Suprofen and the flank syndrome. J Am Med Assoc 259(8):1203–1204

Sjöberg P (1997) Maximum duration of non-rodent toxicity studies: regulatory perspective. In: International conference on harmonization of technical requirements for registration of pharmaceuticals for human use (ICH 4 1997), Proceedings of the fourth international conference on harmonization, Brussels, pp 289–292

Spindler P, Sjöberg P, Knudsen L (2000) First exposure in man: toxicological considerations. Pharmacol Toxicol 86:8–12

Strandberg K (1995) Overview of the outcome of the safety symposium. In: International conference on harmonization of technical requirements for registration of pharmaceuticals for human use (ICH 3 1995), Proceedings of the third international conference on harmonization, Yokohama, pp 554–556

Van Cauteren H, Lumley CE (1997) Harmonization of international toxicity testing guidelines for pharmaceuticals. Contributions to refinement and reduction in animal use. EBRA Bull:4–9

Van Cauteren H, Bentley P, Bode G, Cordier A, Coussment W, Heining P, Sims J (2000) The industry view on long-term toxicology testing in drug development of human pharmaceuticals. Pharmacol Toxicol 86(Suppl I):1–5

Vaughan DP, Tucker GF (1987) Suprofen toxicity. Pharmaceut J 9:577–578

Chapter 9
Why and How Did Reproduction Toxicity Testing Make Its Early Entry into and Rapid Success in ICH?

Rolf Bass, Yasuo Ohno, and Beate Ulbrich

Abstract The need for routine testing of future medicinal products for toxicity to reproduction and development became obvious through the thalidomide disaster. New legal requirements were created, leading to introduction of different testing schemes all over the world. When attempting to fulfil requirements worldwide, somewhat nonsensical piling up of test packages occurred with undue duplication wasting animals.

The ICH topic of reproductive toxicity testing benefited from an already well-oiled machinery of networked experts spanning all ICH parties. Scientifically sound and supported by the six-pack, a draft guideline was presented and accepted, calling for a *most probable option* with three tests focusing on early embryonic development, organogenesis and postnatal development. An apparent dent, how to best accommodate male fertility testing, was also resolved scientifically and agreed by ICH 2.

R. Bass (✉)
Pharmaceutical Medicine, University of Basel, Basel, Switzerland

Pharmacology and Toxicology, Charité, Berlin, Germany

BfArM, Bonn, Germany

Boelckestrasse 80, 12101 Berlin–Tempelhof, Germany
e-mail: rolf_bass@web.de

Y. Ohno
National Institute of Health Sciences, Setagaya, Tokyo, Japan
e-mail: ohno@nihs.go.jp

B. Ulbrich

BfArM, Bonn, Germany

BFR, Berlin, Germany

J.W. van der Laan and J.J. DeGeorge (eds.), *Global Approach in Safety Testing*, AAPS Advances in the Pharmaceutical Sciences Series 5, DOI 10.1007/978-1-4614-5950-7_9,

9.1 Aim of the Guideline

The aim of the so-called tripartite guideline is to support and guide the detection of potential *effects on reproduction* of medicinal product ingredients through scientifically sound experimental investigations in (or without) animals. ICH is the idea of harmonization of hitherto deviating experimental set-ups and secondly to harmonised performance of the testing itself (*the testing*). Harmonization of assessment of study results (*the judgement*) thus obtained in the ICH regions remains of concern beyond creating this tripartite guideline—as for any other *first-generation* ICH guideline.

"Effects on reproduction" must be understood as the smallest common denominator incorporating the detection of toxicity expressed in both parental and offspring generations, and irrespective of the type and severity type of toxic effect exerted. It thus follows that effects on reproduction may become visible at a vast variety of endpoints, which seem extremely difficult to test for in one experimental setting.

Through achieving harmonization, a number of co-aims will be addressed:

- The harmonization of testing *requirements* (*without attempt to impact on high-level regional legislation*): Whereas sponsors before harmonization needed to obey and follow requirements and rules and guidelines established for medicinal products testing in each of the three ICH regions, such—partially deviating—requirements will be eliminated. *NB*: *when ICH started*, *medicinal products testing also varied inside the regions*, *e.g. between the Member States of the European Union*.
- The harmonization of *testing* (*aiming at pharmaceuticals) will impact on other areas*: Harmonization across the pharmaceutical world of active substances needing testing will also support harmonization of testing of chemicals, pesticides and commodities (see OECD). *NB*: *when ICH started*, *medicinal products derived by biotechnological means were only in their early beginnings and not addressed originally*.
- The harmonization of requirements and testing: Offering harmonised guidance beyond the ICH regions (as also requested by WHO). *NB*: *active involvement of WHO introduced Eastern European scientific experts to ICH*. *WHO also introduced participation of and contributions from EFTA States*, *like Switzerland*.
- The harmonization of *judgement*: Appreciating the need for identification of those areas/issues not (yet) harmonised for future harmonization approaches, e.g. assessment of results obtained according to the ICH guideline, and their judgement for risk-benefit balance is, therefore, aimed at detecting the most probable damages to reproduction, and their judgement as being potentially relevant (highly significant), when such medicinal product will finally be used in humans. *NB*: *the deficit of early ICH guideline developments has since been recognised and remedied*.

At the same time, it has to be acknowledged that ICH started off to harmonise experimental execution of testing requirements by sponsors to support best possible

decisions on *marketing authorisation of new medicinal products* by all competent authorities involved in ICH. Requirements for preclinical testing to support *Clinical Development of medicines towards marketing authorisation* were not on the first ICH agenda.

9.2 Introduction

9.2.1 Revolution Without Evolution

Reproductive toxicity testing is the only area in experimental toxicity test settings, where there is a strict *before* (there were no strict testing criteria and conditions set out) and *after* (such testing was now required)—the difference being thalidomide: *revolution*. Other areas of regulatory toxicity testing have developed over time: *evolution*. Things can go wrong during revolution, and so they did for reproductive toxicity testing, with different parts of the world establishing requirements of similar intention with important different details.

9.2.2 Thalidomide Disaster: Before and After the Revolution

Thalidomide had been developed to treat sleeping problems. It established itself to be better, more reliable and non-addictive—compared to barbiturates. At some stage it was the only chemical species available for serious treatment of sleeping disorders and at the same time non-toxic (compared, e.g. to the liver toxicity of barbiturates and their potentially lethal overdosing). This was deemed huge progress, and it was no surprise that thalidomide was soon often recommended also for use in morning sickness and sleeping problems frequently encountered by women during their early pregnancy.

This marketing success was based on an apparent 1950s scientific state-of-the-art industrial development of this substance for use as a medicinal product for human use: medicinal products legislation *pre-thalidomide* did not require reproduction toxicity studies, and such studies were not routinely performed. Thalidomide-containing medicinal products were registered in October 1957 as medicinal specialty in Germany and other European and non-European countries.[1] Soon after, more and more babies were born with typical malformations, frequently concerning the limbs (developing from *limb buds*)—from their complete absence to finger and

[1] It has to be pointed out that during these years, medicines in Germany were not authorised, but registered, meaning a rather passive role for the drug regulatory authority: registration of documents received.

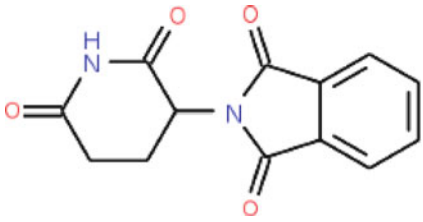


Fig. 9.1 *Chemical structure of thalidomide* (CAS No 50.35-1; MW 258,2), where the two chiral forms +/− freely exchange/revert in vivo in humans

toe malformations, almost anything was observed—depending on the exact time and period of exposure during early pregnancy. Peak occurrence of malformations in Germany was approximately in October 1961 (corresponding with highest Contergan intake ca. 8 months before—i.e. the duration of one pregnancy minus approximately 6 weeks—pointing out the *sensitive period* leading to embryotoxic and teratogenic effects), when Contergan was withdrawn from the market. The structure of thalidomide is given in Fig. 9.1. Action was taken following persisting inquisitive claims made, namely, by two doctors in Germany (Widukind Lenz) and in Australia (William Mc Bride), when malformed babies were seen at one quarter the number at peak in 1961, but already very much above *baseline* frequencies. It must have been terrible to take action and nevertheless to see numbers of affected newborn to go up steeply for the duration of almost *one pregnancy*, when occurrence malformations quickly seized. Case reports had led to suspicion and action and were followed up by epidemiological studies confirming thalidomide to be the cause. Still later, in a *retrospective* attempt to find an animal model that would be able to mimic and confirm the embryotoxic effects occurring in human embryos, it was found that the usual test animal species, namely, mice and rats, were unsuitable to detect the deleterious effects on human fetuses whose mothers had been given thalidomide during a very well-defined *sensitive period* of their pregnancy; rabbits, however, showed somewhat similar malformations albeit after very high doses of thalidomide and other oddities; clearly enough, these observations in rabbits would have led to further questioning and/or prevention of its use during pregnancy in women. These experimental test results disproved the assumption that rodent species would suffice to detect any serious toxicity of a chemical including that to reproduction. NB: also the neurotoxic effects of thalidomide observed in adult humans had not been detected in pre-marketing testing, but this never gained the illustriousness of its teratogenic, embryotoxic including embryolethal effects. The most prominently affected country, where thousands of *Contergan babies* were born, was Germany: children born with typical malformation patterns—ranging from small effects to affection of all extremities. For details of the thalidomide tragedy, its impact and investigations of the mechanism of action, see Jödicke and Neubert (2004). Neubert, when Gruenenthal (the company that developed thalidomide and marketed it as Contergan in Germany) was taken to court, was among the first to seriously investigate the possible mechanism of action leading to the teratogenic effects of thalidomide—and Rolf Bass was one of his students to become a medical doctor in 1965/1967. My own interest in regulating reproductive toxicity

and its testing became clear after my first meeting with Francis Kelsey in 1980 at her FDA office. She was the scientific officer at FDA preventing thalidomide from coming to market in the United States by not approving thalidomide-containing medicines, because she had concerns about the drug's safety, which is a long story in itself.

9.2.3 Details of Damage Impacting on Testing Requirements

Immediately after the causal relationship had been established between therapeutic use of thalidomide during early human pregnancy and the induction of severe malformations, which were very clearly related to a narrow window of susceptibility of embryonic development at an extremely high expression rate, not only pharmaceutical industry but also health ministries and regulatory authorities had to defend themselves to answer not only the accusing question whether this drug disaster could have been prevented but also that of how it should have been prevented. This led to serious changes in drug legislation worldwide, requiring in turn very serious attempts to exclude such potential danger through proper experimental investigations before any marketing authorisation could be granted. The intention was to develop and make available those scientifically sound experimental settings which would allow for addressing and resolving the peculiarities of danger to reproduction through administration of pharmaceutical ingredients, namely, species specificity, substance specificity, dose specificity and toxic effect specificity. Quite understandably, health ministries and authorities all over the world were pressed to introduce appropriate changes somewhat haphazard and without due consideration of *harmonised testing*. As a result, the three future ICH regions and other countries implemented testing for reproductive toxicity, which deviated in general approach as well as in experimental details. Doors were wide open once harmonization was to set in.

9.2.4 There Is No One-Size-Fits-All Approach to Testing

Quite understandably the multiple testing details needed to cover the vast area of potential *effects on reproduction* resulted in somewhat overlapping tests and test schedules. No *one-size-fits-all* experimental approach could cope with the immense difficulties encountered. Guidelines had to be developed in accordance with newly created legal requirements and also needed to be built on science. All developed countries soon had their own ideas about how to prescribe testing for reproductive toxicity—usually addressing active substances to be used in pharmaceuticals/medicinal products. Companies developing such chemical compounds into medicinal products on an international basis soon complained about useless duplication of experimental attempts serving the same purpose without further improving drug safety.

9.2.5 *Early Ideas About Harmonization*

In Europe, for the first time, the Member States of the European Economic Community (EEC) started a harmonised approach, first through legislation (Directive 65/65/EEC, to be modified frequently thereafter) and secondly through development of guidelines (according to Directive 75/318/EEC), recommended for use by sponsors. Such EEC Directives needed national implementation. EFTA countries and the rest of the Western developed world acted similarly but independently. Eastern countries as organised in COMECON for economical reasons set up similar activities for developing their own (or copied versions of) medicines—hoping to export and to make superfluous expensive imports; they were involved in WHO activities looking to protect patients also in third-world countries. NB: please note that EU pharmaceuticals legislation has been changed, enlarged and finally codified under various numbering systems.

9.2.6 *Harmonization of Testing Requirements Deemed Necessary*

Against this background prone to lead to numerous different guideline requirements to be fulfilled by internationally active pharmaceutical companies, various activities evolved aiming at a scientifically sound harmonization. Such activities will be described below. Companies did not want to perform more and more studies that piled up from various countries' requirements, and regulatory authorities had to defend their requirements scientifically. International harmonization was borne.

9.3 Pre-ICH Activities

OECD since long ago has developed guidelines based on long-term exposure of humans (e.g. workers potentially exposed to chemical substances at the work place, often at rather low doses). Animal experimental studies were often constructed to mimic workplace exposure: long-term treatment of animals, often via inhalation and usually at highest possible dose regimes. To address toxicity to reproduction, OECD put in place guidelines for testing of reproductive toxicity prescribing a combination of tests [usually prenatal toxicity studies in rats and rabbits (OECD 414 (1981)), one-generation study (OECD 415 (1983a)) and a two-generation study in rats (OECD 416 (1983b)].

For medicinal products the specific exposure of patients, where treatment for the purpose of *diagnosing* or *healing* is often of limited duration and at rather high doses and always at doses based on pharmacological and pharmacokinetic

knowledge, the proper testing is not always catered for by the OECD approach. Treatment has to ascertain exposure—through suitable routes of exposure (oral, parenteral, local, etc.). Exposure, especially during prenatal development, is searching for *sensitive treatment periods*, unlike, e.g. repeated-dose toxicity testing. Therefore, the experimental scenarios for exposure to active pharmaceutical substances in reproductive toxicity studies have become very specialised. Such has been addressed in the guidelines generated in the European Union, USA, Canada, Japan, etc. following thalidomide.

The European Community set up a task force to describe the scientific basis and need for testing for reproductive toxicity, irrespective of product use and underlying specific legislation. This ground-laying work was published and served as a basis for all future harmonization work (Sullivan et al. 1993). Rolf Bass and Beate Ulbrich were members of this task force.

Before ICH, Rolf Bass had initiated his own *private harmonization activities*. *At first*, the need for revision of the previously adopted guidelines was made obvious at the European Communities Safety Working Party of the CPMP, where Rolf Bass represented Germany and presented my own ideas.

From there Rolf Bass took activities to the United States and Japan. In these two countries, Rolf Bass involved their pharmaceuticals manufacturers' associations (PMA in the USA and JPMA in Japan) as well as their relevant authorities (FDA and EPA in the USA and MHW in Japan) during face-to-face meetings and in subsequent written communication. The European contribution to the outcome, proposed harmonization, was published in 1991 (Draft No. 12 of the evolving international guideline: R. Bass and B. Ulbrich).

Thirdly, the strive for international consensus building involving experts from Japan, Europe and the USA, in the meantime, had progressed towards preparing a rather mature *Draft guideline on detection of toxicity to reproduction for medicinal products* (Bass et al. 1991); this was underwritten by all authors (this draft corresponds to the famous *Draft No. 17* of the evolving international guideline, referred to by many remembering the truly international development process).

Finally, this draft guideline was taken to the scientific community platform representing the worldwide expertise in reproductive toxicity, namely, the Teratology Societies of Australia, Europe, Japan and the USA (International Federation of Teratology Societies—IFTS). At their joint meeting in Sidney in 1990, the state of guideline development was presented, discussed and supported. At the June 1991 IFTS meeting, the IFTS Scientific Program supported further the dissemination and discussion of guideline harmonization activities.

Such dissemination was supported through publication in the International Journal of Toxicology (by the late M. S. Christian, 1992); this took harmonization to yet another level that of consensus of scientific societies worldwide. Thus, both the regulators and the scientists involved in the testing and evaluation of reproductive

toxicity studies agreed in principle on *one* guideline suitable to encompass *all possible testing need scenarios*. This draft guideline was introduced as a well-advanced proposal for harmonised testing requirements into the early ICH activities. The transition from IFTS/SOT to ICH activities occurred at the above-mentioned *Draft No. 17* of a future common guideline and addressed pharmaceuticals only. The ICH guideline evolved rapidly and was presented at ICH 1 in 1991 from all different angles, positions and shortcomings to be overcome by the ICH representatives, and finally adopted 1993 as first *ICH step 3/4 guideline* at the ICH 2 Conference in the USA (ICH, 1993).

For medicinal products the wish and need for harmonization in order to achieve the intended purpose, equal protection of patients anywhere in the world, culminated in the creation of ICH. Rolf Bass became leader of the ICH Safety Topic task force on reproductive toxicity testing. ICH had accepted this topic on the basis that " … none of the (current) testing programs related to the assessment of toxicity to reproduction is materially better or worse under all circumstances than any of the other ones available for the three regions Europe, Japan and North-America" (Bass 1994).

9.4 Towards ICH

9.4.1 Status Quo Ante ICH in the EEC (1)

In relation to Substances contained in Proprietary Medicinal Products

Directive 65/65/EEC for the first time introduced requirements to document quality, safety and efficacy within an application for marketing authorization. This requirement resulted from the thalidomide catastrophe. Among other requirements:

Article 4 states that:

> The application shall be accompanied by the following particulars and documents: 5. Results of: …—pharmacological and toxicological tests …so that the procedure for granting a marketing authorization may be completed as prescribed.

In Article 4b it is stated that:

> Furthermore, the competent authorities shall draw up an assessment report and comments on the dossier as regards the results of the analytical and pharmaco-/toxicological tests and the clinical trials of the medicinal product concerned. The assessment report shall be updated whenever new information becomes available which is of importance for the evaluation of the quality, safety or efficacy of the medicinal product concerned.

This is further detailed in *Directive 75/318/EEC*:

> … that applications for authorization to place a proprietary medicinal product on the market should be accompanied by particulars and documents relating to the results of tests and trials carried out on the product concerned;

Annex to this Directive sets out in:

PART 2 TOXICOLOGICAL AND PHARMACOLOGICAL TESTS

The particulars and documents accompanying the application for marketing authorization pursuant to point 8 of Article 4, second paragraph, of Directive 65/65/EEC shall be given in accordance with the requirements of Chapters I and II below.

CHAPTER I PERFORMANCE OF TESTS

A. INTRODUCTION

The toxicological and pharmacological tests must show: 1. the potential toxicity of the product and any dangerous or undesirable toxic effects that may occur under the proposed conditions of use in human beings; these should be evaluated in relation to the gravity of the pathological condition concerned; … All results must be reliable and of general applicability …

Under B: TOXICITY

C. FOETAL TOXICITY

This investigation comprises a demonstration of the toxic and especially the teratogenic effects observed in the issue of conception when the substance under investigation has been administered to the female during pregnancy. Although up to the present these tests have had only a limited predictive value in regard to the application of the results to human beings, they are thought to provide important information where the results show effects such as resorptions and other anomalies. … The tests in question shall be carried out on at least two animal species: a breed of rabbits sensitive to known teratogenic substances and rats or mice (specifying the strain) or, if appropriate, in some other animal species. The details of the test (number of animals, amounts administered, timing of administration and criteria for evaluation of results) shall depend on the state of scientific knowledge at the time when the application is lodged, and the level of statistical significance that the results must attain.

D. EXAMINATION OF REPRODUCTIVE FUNCTION

If the results of other tests reveal anything suggesting harmful effects on progeny or impairment of male or female reproductive function, this shall be investigated by appropriate tests.

These, according to European implementation procedures, needed to be transferred into national pharmaceutical legislation and administrative procedures in each Member State. Such led to requiring performance of studies fulfilling the requirements as laid out—usually before and during clinical trials (which remained under pure national control)—and their documentation for submission within the application for obtaining a marketing authorization in Member States of the EEC. These studies became known as segments 1, 2 and 3 studies addressing all of the issues stated above. Such studies had been introduced also in Japan and the USA—the ICH partners.

Administrative and regulatory details were provided in Directive 75/319/EEC, requiring expertise not only for the performance of studies but also for their description and documentation.

Subsequently, Notes for Guidance were developed by the Safety Working Party of the CPMP and adopted by the CPMP for implementation by the European

Commission (and first published by the European Commission in Volume 3 of their legislation—Eudralex).[2]

These Notes for Guidance had to be developed according to the post-thalidomide pharmaceutical legislation as set out (Directive 75/318/EEC). This meant that for the European Community, guidelines had to be written addressing *foetal toxicity* and *reproductive function* in separate guidelines. Although the scientific shortcomings of these splitting parts of the same, i.e. overall *effects on reproduction* toxicity testing, were obvious, legislation had to be followed and was implemented correctly. These two Notes for Guidance were applied by sponsors developing proprietary medicinal products for submission for obtaining marketing authorisation in Member States of the EEC and by the regulatory authorities to review the usefulness and sufficiency of studies and data for obtaining such marketing authorization and under which conditions in relation to reproductive toxicity. Application and use made clear that the shortcomings predicted existed also in practice, and improvements were deemed necessary. Similar experience developed in Japan and the USA.

The European Notes for Guidance were presented and discussed concerning their usefulness first during official bilateral discussions between the European Commission and the USA and Japan—pre-ICH. The preclinical discussions held showed equal interest of the US and Japanese regulators to revise reproductive toxicity testing requirements. When ICH was created, the topic of reproductive toxicity testing was born naturally. The European contribution to ICH (Bass and Ulbrich 1991) was used together with Draft No. 17 of the IFTS activities, to establish *Topic ICH Safety 5* to achieve harmonization of testing requirements in the area of reproductive toxicity testing.[3]

9.4.2 Status Quo Ante ICH in the EEC (2)

In relation to the toxicology of chemicals, the Commission of the European Communities, Directorate General for Employment, Industrial Relations and Social Affairs—Health and Safety Directorate, Luxembourg, saw the need to develop guidelines for the evaluation of reproductive toxicity. For this purpose, Marie-Therese van der Venne together with Alexander Berlin from the European

[2] Rolf Bass had the pleasure to serve on the Safety Working Party of the CPMP as member for Germany during the finalisation of the guidelines. Later on, as Chairman of the Working Party, Rolf Bass took on responsibility for the further development and international positioning of these guidelines.

[3] Rolf Bass had the pleasure to first work with John Griffin and then to take over full responsibility for non-clinical testing during the discussions between the European Commission and the USA and Japan. The European delegation, led by Fermand Sauer from the European Commission, included Anthony Cartright (from the UK authority) who addressed pharmaceutical quality; Jean-Michel Alexandre (from the French authority), who addressed efficacy; and Rolf Bass (from the German authority) addressing preclinical safety. As we remember, these bilateral discussions made obvious the need for international harmonization of non-clinical testing requirements to be fulfilled for obtaining marketing authorization.

Commission established an ad hoc working group—guided by Diether Neubert—which met from 1985 onwards. Their activities led to the publication of a monograph in 1993 (Guidelines for the evaluation of reproductive toxicity: F.M. Sullivan et al., eds). Interestingly enough:

> … it was the task of the working group to describe the state of the art of how to assess all aspects of the reproductive toxicity of chemical substances. The assessment is to be made independently of the source or area of commercial use of the chemical substance to be investigated. The description on how to assess reproductive toxicity, therefore, should cover various different types of chemical substances and their prospective use.

Nota bene, this scientific guideline thus addressed not only the testing and assessment of *chemicals* but also those active substances to be included in proprietary medicinal products, but was not bound by the definitions and restrictions of Directive 75/318/EEC. The guideline will be addressed and described in more detail below.[4]

9.4.3 *Status Quo Ante ICH in the EEC (3)* *(Wording and Content Extracted from Sullivan et al. 1993)*

The European ad hoc working group developed guidelines for the testing and evaluation of reproductive toxicity based on the fact that reproduction is concerned with the continuation of the species. The reproductive process is basically similar in all mammals, but in each species there are subtle variations complementary to the species ecological niche. This can be taken as a starting point to foresee the many difficulties in mimicking the occurrence of toxic effects on reproduction across species and in a priori assuming similar effects—be it occurrence of toxicity

[4] Together with Beate Ulbrich, Rolf Bass had the pleasure to represent the Drug Institute of the BGA on this ad hoc working group. The outcome, prior to its publication, was used to initiate discussions with the USA and Japan (see above), where the intention was to start harmonization of reproductive toxicity testing mainly for medicinal products. Rolf Bass was in contact with and travelled to the US-FDA and the US Pharmaceuticals manufacturers association (then PhRMA), and the Japanese Ministry of Health (in charge of handling medicinal products)—here the late Professor Yoshihito Omori (Omori 1991) must be mentioned for his openness and scientific strictness towards the regulation of medicines and their toxicological testing as well as for his personal kindness towards the newcomer to Japanese culture—and the Japanese Pharmaceutical Manufacturers Association (JPMA). The very positive outcomes from these discussions helped me to take the next step: discussion with the European Teratology Society. This quickly led to approaching the International Federation of Teratology Societies (IFTS), where a Committee on International Regulation was designated to interact with the recent European developments. On occasion of the IFTS meeting in Sidney, Christmas 1990, where all interested parties were present, Rolf Bass presented the current development. IFTS welcomed the harmonization approach (see publication by Christian 1992). Already in the preparation for ICH 1, the topic of reproductive toxicity testing was tabled, where Rolf Bass became the lead for the topic Safety 5—Reproductive Toxicity Testing. In November 1991, at ICH 1 in Brussels, the draft guideline was presented at step 3 of the ICH process (for further details, see below).

or apparent non-toxicity—to occur in the tested animal species and the target—humans. In addition, there are numerous differences in handling substances which are species-specific, and these further contribute to the sometimes immense difficulties of extrapolating effects across species; such is known not only for reproductive toxicity effects.

Reproduction includes a sequence of events, which for better understanding can be presented as a cycle, where toxicity can be introduced at any point, and toxicity can become visible at any point.

One may start with:

(a) premating to conception, followed by
(b) conception to implantation, then
(c) implantation to palatal closure, and then
(d) palatal c losure to term of pregnancy/birth, then
(e) birth to weaning, and finally
(f) weaning to sexual maturity sexual maturation,

Where all steps concern more than one generation (see also Table 4 Testing strategy: Bass 1991).

This makes understandable the wish for exposing throughout all steps of reproduction and to assess potential effects whenever possible. For feasibility reasons, both treatment periods and points for assessing toxic effects need to be reduced to as few as possible combinations, which can be addressed experimentally, and which relate to exposure—active substances of medicinal products. Thus, the reproductive cycle has been broken up into three periods, which are thought to be representative for possible medical treatment periods, especially sensitive to such treatment, and provide for accessible endpoints. These are the now well-known periods of treatment, which can be assessed as *fertility* (segment 1: effects on the reproductive system), from *implantation to organogenesis* (segment 2: embryo–fetal toxicity) and *fetal to postnatal development* (segment 3: peri-postnatal toxicity). Maternal toxicity impacting on reproduction needs to be kept in mind as well. It is quite clear that these may look arbitrary and not suitable to represent certain exposures to medicinal products under real treatment conditions. Therefore, exposure throughout one complete cycle or even longer and assessment of various endpoints in one and the same experiment constitutes another approach (seldom used for pharmaceuticals). Overall, application of an insult at any point in this integrated process may have far-reaching consequences for successive generations. The disastrous example of thalidomide let us focus on that portion of reproduction, which induces damages during organogenesis, but routine investigations of new medicinal compounds must also address in arguing and/or experimentation the other exposure periods described above. All stages of reproduction are vulnerable to a lesser or greater extent, and the effects may be induced directly or indirectly. Periods of rapid developmental changes may be more susceptible or vulnerable than other periods, but there is no such exclusivity. Selecting studies and combinations of studies that are sensible depends on the approach taken, the preferences of the investigator and the needs of the regulator, as depicted in Fig. 9.2.

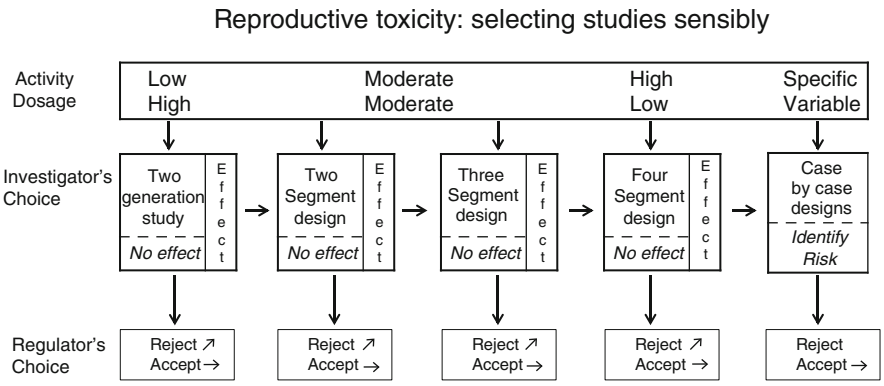


Fig. 9.2 *Reproductive toxicity studies: selecting studies sensibly*. Activity and dosage expected as therapeutic use are given as two factors discriminating dosing periods for the studies. Most often the *most preferred option* (see text) will be chosen by the investigator. As there are many other factors which possibly influence the experimental design, these *specific variables* may lead to completely different designs to identify risk to human pregnancy satisfying both the investigator and the regulator (From R. Bass, Proceedings of ICH 1, Figure 2, p. 288)

Standardised tests, testing and assessment of effects observed in the field of reproductive toxicity are among the most difficult areas of experimental toxicity. Even a precise definition of what constitutes *reproductive toxicity* remains difficult, as the same type of response may be desirable (i.e. pharmacological) or undesirable (toxic)—according to circumstance. As can be seen for various reproductive toxicants, effects on reproduction can predominate or occur at exposures considerably lower than those causing other forms of (adult/parental) toxicity or may be hidden due to dominating parental toxicity. This hints towards testing over the full range of exposures which are likely to occur in the reality of therapeutic treatment.

Whereas all stages of reproduction are vulnerable, some stages are more susceptible than others. Thus, reproductive toxicity testing has often been mistakenly understood as *teratogenicity*, which would falsely reduce the wide range of toxicity to visible malformations, a conceivable but wrong approach. Common and frequent reaction to early interference with the normal development process is prenatal wastage (and also postnatal death). In humans, about 2–3% of newborns exhibit major structural abnormalities (in 1% they may be so pronounced that without surgery they are not compatible with life). Eight percent of children at the age of 5 years show signs of either morphological or functional disorders. Types of adverse effects to be ascertained in experimental testing address the following times and targets at which a substance initiates its toxicity:

- Adult toxicity (effects, e.g. on libido, mating, gamete production)
- Maternal toxicity (effects, e.g. on ability to nurse)

- Developmental toxicity, split into (and addressed in regular reproductive toxicity testing approaches):
  - Preimplantation and implantation (effects, e.g. on fertilisation, implantation)
  - Embryonic development (effects, e.g. on organ development, growth and survival)
  - Placental development (effects, e.g. on growth and organ function)
  - Fetal development (effects, e.g. on growth, differentiation and development, survival)
  - Postnatal development (effects, e.g. on hormone and immune function, CNS and peripheral NS function, sexual function, survival)

With this in mind, it is obvious that case-by-case requirements and interpretation of data thus generated are quite dependent on the personal knowledge and expertise of the sponsor's experts and the authorities' assessors. Such has been known for long to create different performances of tests and the judgement of their results. As described for Europe (and certainly similarly true in Japan and North America) as status quo ante above, the sequence of pre-ICH events and activities gave this topic a jump-start towards early success and presentation of a harmonised ICH guideline for reproductive toxicity testing, which went well beyond its regional predecessors allowing for state-of-the-art testing.

9.4.4 Status Quo Ante ICH in Japan

The Japanese guideline on reproduction toxicity was prepared after the thalidomide tragedy. It differed from those of the other countries in period of drug administration, highest dose to be administered, some parameters for evaluation, etc. and caused a problem hindering the acceptance of foreign data. Therefore, the Japanese Ministry of Health and Welfare (now, Ministry of Health, Labour and Welfare) initiated research on the influence of those differences on the detection of compounds with reproductive toxicity. It was finally notified that the differences in the dosing period and the highest dose between Japan, the USA, English, EC and Canada were of minor importance, and that the data obtained according to the recommendation of those countries can be accepted (MHLW 1989). However, it remained difficult to accept data in Japan, in which the requirements on some parameters were not satisfied. Especially those of rodent segment II tests differed significantly. Similarly, a problem existed, hindering acceptance of Japanese data in foreign countries. Therefore, Japan considered it necessary to produce internationally harmonised guidelines for reproduction tests and Japanese experts cooperated with representatives of the EU and USA in the framework of ICH.

9.5 Reproductive Toxicity Testing as ICH Activity S5

As for any other ICH activity, the topic of a future harmonised guideline needed agreement at Steering Committee level. The *Joint ICH Expert Working Group* (*EWG*) *on Safety* then took on the task of actually achieving such harmonization. The working group included participants from the usual ICH partners. The term *six-pack* was introduced by our safety group, which had the backing of the partners involved in the previous activities described above. All regional regulators and pharmaceutical industries associations had already been involved, and the six-pack could start on the harmonization achieved at IFTS level. The official experts signing off the step 3 guideline in June 1993 as sponsors were Alan Taylor (succeeding Judy Weissinger) and Joy Cavagnaro for FDA's CDER and CBER, Yuzo Hayashi for MHW and Rolf Bass for EC. Following presentation to the Steering Committee, the EWG claimed to have achieved harmonization of the first guideline (official implementation by the three regional authorities to be awaited).

Harmonization achieved at such scientific level needed transposition into an ICH guideline. Albeit striving for harmonization, such ICH tripartite guideline was not allowed to contradict the three regional legislations active at that time (e.g. requiring separate testing requirements for *fetal toxicity* and *examination of reproductive function* in the EC, and other peculiarities in the USA and Japan). The Steering Committee would and could not allow any guideline development, which would require one or more of the regions to change their drug legislation.

Here, the *segmental approach* to reproductive toxicity testing needs to be set out. Figures 9.3, 9.4 and 9.5 describe in principle segments 1, 2 and 3 as treatment periods versus time points of assessment (notwithstanding any modification due to experimental design, animal species and pharmacokinetic behaviour of the test substance). Examples are given for various treatment periods: in black the *preferred option* favoured in the ICH guideline is given; in red and yellow, other, arbitrarily longer, treatment periods are sketched to show the flexibility of the segmental treatment approach—according to the knowledge about the substance and the need to know. It is obvious even to the nonexpert in reproductive toxicity testing that regulators' hobby horses, substance peculiarities and other issues can easily widen or shorten the span of each segmental treatment period, or even split them into shorter ones (e.g. to focus on one of the typical thalidomide malformations), and can favour higher or lower dosage regimes, thus resulting in minute or coarse deviation from the preferred and harmonised approach. There must be scientific understanding of what is needed (as experimental result or endpoint) versus the design (experimental power) to be chosen. Although the concept of target organ toxicity is true also for reproduction toxicity, it rather involves broad spectra of potential sites and experimental endpoints (and mechanism of action) compared to more narrow potential sites and experimental endpoints as visualised in mutagenesis and carcinogenesis.

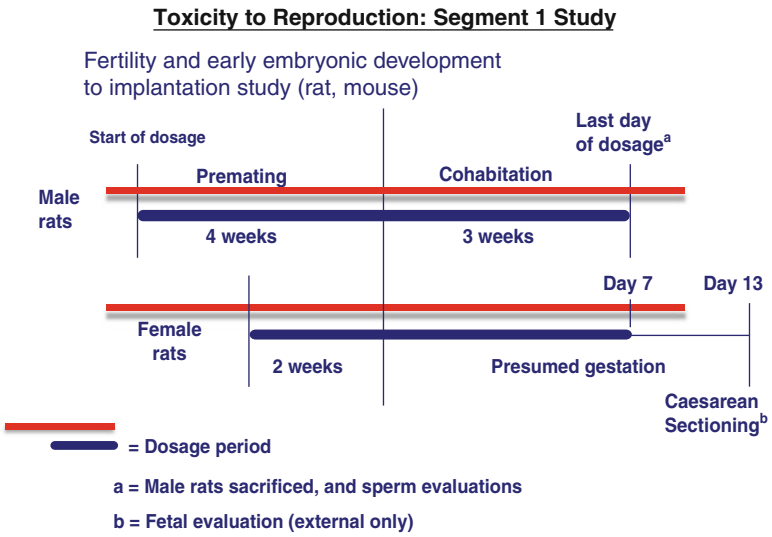


Fig. 9.3 *Toxicity to reproduction* (*segment 1 study*). For the *preferred option* (see text), treatment of male and female animals (rats) would last from premating until day 7 of presumed gestation (females) (*blue colour*). Caesarian sectioning to take place on day 13 of presumed gestation. Treatment periods can be prolonged (*red colour*) depending on the conditions depicted in Fig. 9.2 [Modified from Christian (2001) and Peters and Garbis-Berkvens (1996)]

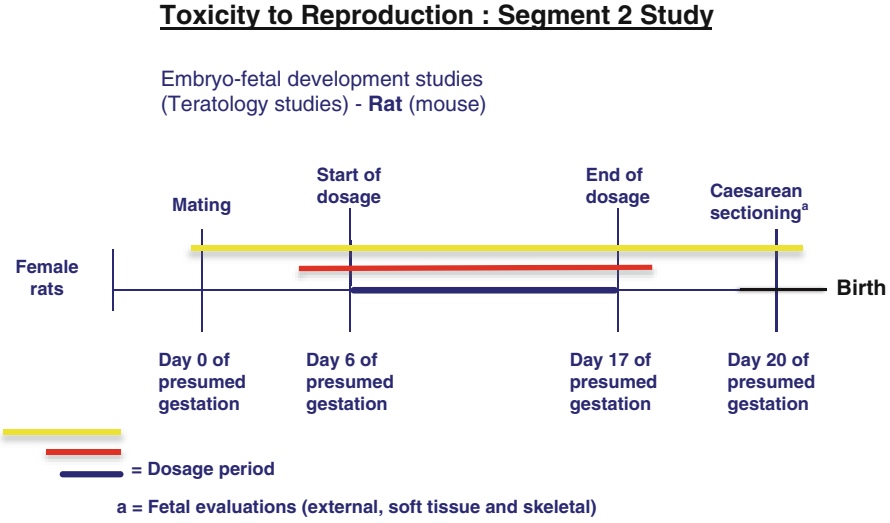


Fig. 9.4 *Toxicity to reproduction* (*segment 2 study*). For the *preferred option* (see text), treatment of female animals (rats or mice) would last from day 7 of presumed gestation to day 17 (*blue colour*) or longer (*red colour* somewhat extended, *yellow colour* also covering the mating period). Caesarian sectioning to take place on day 20 of presumed gestation. For other animal species (rabbits, non-human primates), treatment periods need to be adapted accordingly. Treatment periods should be chosen to accommodate the conditions depicted in Fig. 9.2 [Modified from Christian (2001) and Peters and Garbis-Berkvens (1996)]

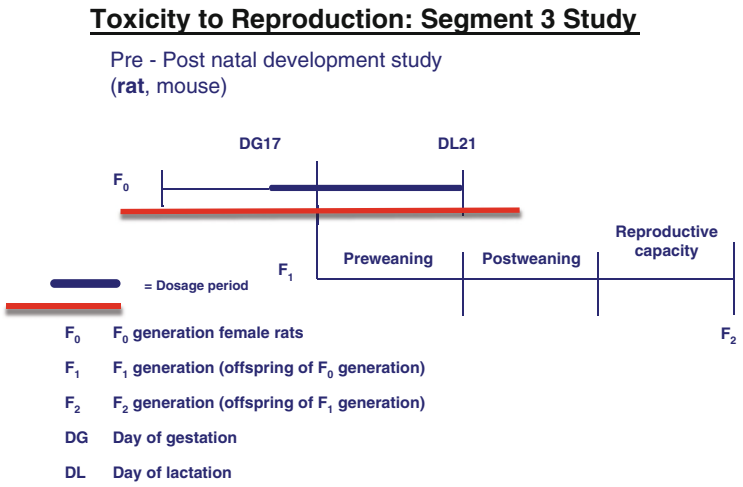


Fig. 9.5 *Toxicity to reproduction* (*segment 3 study*). For the *preferred option* (see text), treatment of female animals (rats or mice—F0 generation) would start during organogenesis to last until normal birth (*blue colour*), or start may be extended (*red colour*)—and possibly last until weanling; F1 generation to be investigated and followed up—possibly to also assess their reproductive capacity—F2 generation. Treatment periods should be chosen to accommodate the conditions depicted in Fig. 9.2 [Modified from Christian (2001) and Peters and Garbis-Berkvens (1996)]

9.5.1 Segment 1 Study (Fig. 9.3)

For these studies the test substance is administered *prior to and in the early stages of pregnancy* (Japanese wording) in order to assess *fertility* (US wording) or *fertility and general reproductive performance* (European wording). These differences do not look very impressive, but they are connected with different duration of treatment: *throughout pregnancy and lactation* (USA), *throughout pregnancy* (EEC) and *to gestation day 7 in rat or day 6 in mice* (Japan). It does not come as a surprise that the endpoints (dates for examination of which parameters) also differed. Whereas in the USA and EEC, half of the females were to be killed at mid-pregnancy, for Japan all animals were assessed at term. This impacts on the parameters, which can be assessed. For Japan on top of testing fertility and successful pregnancy, also growth and development of the fetuses had to be investigated, which were to be addressed in the US and EEC studies in a different manner. Harmonization of this study proved most difficult. The final goal of including segment 1 in a repeat-dose study was to be achieved, but only later (of course, suspicious results obtained in such repeat-dose study would trigger more specific investigations resembling either of the pre-ICH regional specifications).

9.5.2 Segment 2 Study (Fig. 9.4)

Despite somewhat different wording as titles for this study type in the USA, Japan and EEC with the exception of number of animals to be used, there were no major differences. Also the endpoints to be assessed where quite similar, but for Japan the offspring (F1 generation) had to be assessed postnatally, which was not included in US and EEC studies. This study is important in order to find and assess damage to the embryo and fetus in all three regions.

9.5.3 Segment 3 Study (Fig. 9.5)

In the titles for this study, the *peri-/postnatal periods* (USA) of development of the offspring are addressed (*perinatal* for EEC and *perinatal and lactation* for Japan), signalling very similar intentions. Despite the different wording for the treatment periods, they very much resembled each other, but numbers of animals to be used were different. With regard to endpoints to be addressed, Japan and EEC had included more postnatal testing up to and including reproductive performance of the offspring. On the other hand, the USA (National Centre for Research of Toxicology—NCTR) had initiated a collaborative study to better understand behavioural testing and its value during the postnatal period (NB: Beate Ulbrich's laboratory at the German authority was included as the only non-US site for this study). This activity shows that all three regions had similar intentions for the segment 3 study, which easily surfaced in the harmonised ICH guideline. Surprisingly, the multitude of test methods and vast number of endpoints developed to better understand normal postnatal development and deviations from it did not hinder the positive attitude; it rather sped up harmonization.

9.5.4 Across and Beyond Segments Studies

It is quite clear that performance of three segments (according to whichever regions current requirements) may or may not suffice to redress issues. Different segmental designs may become necessary; treatment periods different from those described for the segments may become necessary. Studies outside the reproductive and developmental testing areas may become necessary for bridging data/effects, or for bridging understanding between species, namely, kinetic and dynamic, investigations may be needed. There is ample opportunity for doing such studies, but only when scientifically justifiable; they cannot be described in any detail here (many of them can be found in the ICH guideline S5). Combinations of currently existing

(regional) regulatory protocols are proposed that would offer complete evaluation and mechanisms to address scientific considerations that are not consistent across the regions (Lumley 1991). Such combinations cannot reliably be predicted and laid down in guideline texts; they must arise on the occasion.

9.5.5 ICH Step-by-Step Approach

Pharmaceutical associations had stressed before that the pre-existing regional requirements had often led to serious duplication of similar experimental set-ups, like EC obeying segments 1, 2 and 3 needed to be complemented by Japanese segments; Japanese obeying segments 1, 2 and 3 needed to be complemented by EC and US segment 1 and by EC segments 2 and 3; and US obeying segments 1, 2 and 3 needed to be complemented by Japanese segments 1 and 2—certainly spelling out disaster for anybody trying to seriously plan animal experiments for reproductive toxicity testing (for details, see Sect. 5.5.3); regulators had to admit the existence of such discrepancies. Duplication of several *second-best* programmes stemming from the three regional guidelines should be replaced by an optimal choice of test design according to the test substance, its peculiarities and knowledge accumulated about the substance over its development (at that point in time, when decisions about testing were to be made). Quite understandably flexibility needed to be incorporated into the testing strategy underlying harmonised guidelines (already at this early stage asking for a *core* battery of tests that could support enrolment of women in clinical trials, not subject of early ICH activities).

There is sufficient knowledge of and confidence in the current practice of the three regions for regulatory authorities to accept studies based on their scientific merit, and carried out according to the requirements of another region. Such flexible attitude will help to bridge the gap between current requirements and future harmonised guidelines.

Not surprisingly, in the process of ICH guideline development, regional industries tended to defend requirements specific for *their* region, and regional arguments offered by regulators fell back onto previous positions. Nevertheless, the six-pack agreed to agree, with some words of caution and disclaimers. These needed to be addressed in the fine-tuning of the flexible approach to testing—often resulting in testing three segments separately. Surprisingly, the three segments proposed to cover many test substance oriented requirements became acceptable, almost. Already at this stage it was clear that many of the previous requirements leading to duplication of tests could be struck out.

A tremendous step forward was the understanding that scientific justification of study designs should be used, and a precise determination, which array of questions each testing procedure (e.g. segment 1, 2 or 3 tests) *can* and *should* answer.

At the end of the ICH 1 Conference it was stated that the differences in current guidelines in the three regions could result in duplication and major redundancies in studies, without contributing to safety. It was, therefore, recommended to quickly proceed towards provision and adoption of harmonised guidelines. In the meantime authorities should open up and mutually accept studies case by case even if not performed according to their regional requirements.

9.5.6 ICH Harmonised Guideline

- From *mutual acceptance* of studies to *harmonised* studies.
  At *ICH 1*, S. Takayama stated that results of studies carried out following the FDA or Japanese guidelines are usually accepted in EC, whereas in the USA and even more frequently in Japan, studies had to be repeated (S. Takayama, ICH 1). From this he developed his ideas to further mutual acceptance based on the following facts and opinions, which were all taken up in the ICH process towards ICH 2:

  1. The guidelines of EC, FDA and Japan include *all* of reproductive and developmental processes as a whole (albeit with somewhat different distribution into segments).
  2. From almost 500 drugs marketed from 1981 to 1990 in the world and from the less than 10% showing occasional side effects on reproductive and developmental function, the correlation of animal data predicting human fertility is poor. Safety of drugs potentially administered during human pregnancy arises from not using them.
  3. Sufficient protection of humans arises from over-predicting human teratogenicity from animal data (thus excluding them from use during pregnancy in humans). *For numbers 4 and 5*, *see below*.

  This shows where mutual acceptance had to set in. The various activities preceding ICH and the common wish to achieve harmonization of guidelines supported speeding up to move from more and more practised mutual acceptance of studies/core packages of studies performed outside one's own region and also performed according to guidelines other than those enacted at home towards agreement of new common and harmonised guidelines, which would embrace approaches from all three regions. Ambitious presentations were given at ICH 1 in 1991 (Takayama 1991; Weissinger 1991; Diener 1991; Bass 1991; van Cauteren 1991—all in ICH 1), and they culminated towards ICH 2, with the agreed and signed achievement of a harmonised guideline for reproductive toxicity testing, which could be presented to the ICH Steering Committee for accepting step 3 of the ICH procedures for further processing towards step 4 (R. Bass, ICH2).

- *From harmonised studies to harmonization of test packages submitted to obtain marketing authorization in the three regions*.

  Now that the *scientific rational* had been introduced in order to arrive at a harmonised guideline, pharmaceutical industry wanted to reap the benefit from applying the guideline, i.e. putting the agreed guideline to the test: acceptance of studies performed according to the guideline as being sufficient to support an application for marketing authorization—notwithstanding assessment of results to document sufficiently *nonreproductive toxic potential*—to the extent that the pharmaceutical product can be used safely during human pregnancy. No, as human data were still lacking at this point in time, drugs usually were not recommended for use in pregnancy, rather their use would be discouraged. Marketing authorisation, therefore, did not include therapeutic use in relation to pregnancy, but would recommend *soft* or *strict* contraindications. Both in the USA and the EEC, this triggered further developments to allow *classification* of medicines from *may be* to *must not* be used; these activities remain outside the ICH developments. The idea of classification, however, can be found in ICH, where later on guidelines on requirements for clinical trials were developed. Here, decisions on the inclusion of women of childbearing potential and those being pregnant or lactating rather early during clinical development had to be taken—based on some studies addressing reproductive toxicity.

Concerning *male and female fertility testing*, the Japanese colleagues pointed out that from currently available data, it is clear that approximately 20% of male fertility impairment cases may not be detected in repeat-dose studies. Although none of the six substances concerned produced demonstrable reproductive adverse effects in male human beings, further discussion was deemed necessary before accepting repeat-dose studies to embrace detection of male fertility impairment. Japan showed keen interest to resolve this issue via further investigation. Thus, Takayama et al. (1995) conducted a study to better understand the validity of 4 weeks dosing in rat as a test to detect male fertility disorders and concluded that a 4-week treatment period is appropriate to detect drug effects on male fertility, if the examination was combined with detailed morphological studies including stage and semen analysis. According to these data, the ICH harmonised guideline was revised in 1995 as S5(R1).

On the other hand, it became apparent during the process of discussion on the timing of non-clinical tests in relation to the clinical trials that there were several differences between three regions, among them the field of reproduction tests. These include toxicity studies to support first entry into man and the recommendations to perform reproduction toxicity studies for women of childbearing potential. We, experts from regulators and industry from the three ICH regions, promised to work towards further harmonization of the ICH-M3 guideline (1997). Because Japanese experts were not convinced that it is possible to sufficiently evaluate male and female fertility organs during a 2-week repeat-dose toxicity

study, Japan had recommended that a 4-week toxicity study should be conducted before starting clinical tests in Japan. After the agreement of the first M3 guidelines, Japan started two separate validations to confirm the validity of 2-week repeat-dose toxicity study.

The first study was *Collaborative work to evaluate toxicity on male reproductive organs by* 2-*week repeat-dose toxicity studies in rats* using a variety of pharmaceuticals and chemicals (24 substances and 30 different protocols) with known toxic effects on male reproductive organs (Table 9.1). In the collaborative study, 2- and 4-week repeat-dose toxicity studies with detailed histopathological examinations of male reproductive organs were conducted. The study was completed in 2000. It can be concluded that a 2-week study with an appropriate dose setting had an equal potential compared to 4-week studies towards detecting the toxicity on male reproductive organs—if precise stage analysis of sperm was also conducted (Sakai et al. 2000). Subsequently, Japan decided to agree with the performance of clinical studies up to phase II without separate male fertility studies, if the evaluation of male reproductive organs was performed in repeat-dose toxicity studies of more than 2-week duration, which was already included in the addendum to the ICH S5 guideline: S5a - *Toxicity to male fertility*, which was accepted at step 4 in 1995 [undergoing minor revisions in 2000: S5(R1) and S5(R2)]. Finally, in 2005, the guideline was revised and renamed: *Detection of toxicity for reproduction of medicinal products and toxicity to male fertility* (S5(R2)). It shows the impact of ICH on achieving and performing: male fertility can now be tested in the *most probable option* via inclusion into a repeat-dose toxicity study.

The second study was *Collaborative work on evaluation of ovarian toxicity by repeat-dose and fertility studies in female rats* using 17 chemicals with known toxic effects on female reproductive organs. It was completed in 2009, and the results indicated that ovarian toxicity could be detected by a careful histopathological examination (Table 9.2; Sanbuissho et al. 2009). A 2-week dosing period may usually be sufficient for the induction of ovarian toxicity, except for cytotoxic compounds such as alkylating agents. These results were reflected to the deletion of the Japanese requirement of female fertility studies before inclusion of woman of childbearing potential up to phase II clinical studies in ICH-M3(R2) guidelines (2009).

J. Weissinger at ICH 1 pointed out that there are six areas significantly different among the approaches to evaluating reproductive and development toxicity in the three regions (Weissinger 1991) (1) day of sacrifice for segment 1 studies, (2) appropriate level of maternal/paternal toxicity that must be observed to understand when an *animal has* (*only*) *been made ill*, (3) possibility of missing findings by not dosing across segments, (4) duration of premating treatment for male fertility testing, (5) impact of kinetic and metabolic data on study design and (6) specification of scenarios when behavioural toxicity studies need to be performed.

Table 9.1 Summary of collaborative work

Compound	Route	Dose levels (mg/kg/day)	Animal strain	No. of animals/ group	Results	Evaluation	Company	Reference
(1) *Hormonal drugs and their antagonists*								
Estradiol benzoate	s.c.	5, 20, 50 and 100 μg/kg for 2 and 4 weeks	Slc:SD	6	Decreases in the reproductive organ weights, degeneration/necrosis of pachytene spermatocytes and atrophy of Leydig cells in the 2- and 4-week groups	Possible by a 2-week study	Teijin	Hata et al.
Ethinylestradiol	p.o.	3 and 10 for 2 weeks and 3 for 4 weeks	Crj:CD (SD)	6	Decreases in the reproductive organ weights in the 2- and 4-week groups. Apoptosis of round spermatids and atrophy of seminiferous tubules in the 4-week group and retention of spermatids in the 2-week groups	Possible by a 2-week study	Toray Industries	Miyamoto et al.
Ethinylestradiol	s.c.	0.3 and 3.0 for 2 and 4 weeks	Crj:CD (SD)	6	Decreases in the reproductive organ weights and atrophy of Leydig cells and degeneration/necrosis of pachytene spermatocytes in the 2- and 4-week groups	Possible by a 2-week study	Zeria Pharmaceutical Co.	Kinomoto et al.

(continued)

Table 9.1 (continued)

Compound	Route	Dose levels (mg/kg/day)	Animal strain	No. of animals/ group	Results	Evaluation	Company	Reference
Fadrozole	p.o.	30 and 60 for 2 and 4 weeks	HanIbm: WIST	6	Decreased weights of seminal vesicles, prostate and epididymides, but not testes. Degeneration/ necrosis of the pachytene spermatocytes was observed, but no changes in sperm examinations in the 2- and 4-week groups	Possible by a 2-week study	Novartis Pharma	Kawashita et al.
Flutamide	p.o.	60 and 200 for 2 and 4 weeks	Crj:CD (SD)	6	Decreases in the reproductive organ weights and Leydig cell proliferation. Number and motility of sperm were decreased in the 2- and 4-week groups	Possible by a 2-week study	Santen Pharmaceutical Co.	Okahara et al.
(2) *Central hormonal modulators*								
Haloperidol	p.o.	30, 60 and 80 for 2 weeks and 30 and 60 for 4 weeks	Crj:CD (SD)	6	Decreases in the reproductive organ weights, atrophy of Leydig cells and necrotic pachytene spermatocytes were observed in the 2- and 4-week groups	Possible by a 2-week study	WelFide Corporation	Kohge et al.

Reserpine	s.c.	0.05, 0.1 and 0.2 for 2 weeks and 0.05 and 0.1 for 4 weeks	Crj:CD (SD)	6	Sperm number and reproductive organ weights were not affected. Histopathologically, retention of step 19 spermatids in the 4-week group and apoptosis of spermatocytes in the 2-week group were observed	Possible by a 2-week study	Fujisawa Pharmaceutical Co.	Yamauchi et al.
Reserpine	s.c.	0.05 and 0.1 for 2 and 4 weeks	Crj:CD (SD) IGS	4–12	Prostate weights were decreased in the 4-week group. Histopathologically, retention of step 19 spermatids was observed in the 2- and 4-week groups, including the control group	No clear lesions in either 2- or 4-week groups	Aventis Pharma	Kawamura et al.
(3) *Nucleic acid modulators*								
Adriamycin	i.v.	1 and 2 mg/kg/week for 2 weeks and 1 mg/kg/week for 4 weeks	Slc:SD	8	Testis weights were decreased and histological examination revealed a decrease in spermatogonia in the 2- and 4-week groups. Stage analysis revealed significantly decreased numbers of spermatogonia in the 2-week group	Possible by a 2-week study	Tanabe Seiyaku Co.	Adachi et al.

(continued)

Table 9.1 (continued)

Compound	Route	Dose levels (mg/kg/day)	Animal strain	No. of animals/ group	Results	Evaluation	Company	Reference
Adriamycin	i.v.	Single dose of 2 and 6 mg/kg	Crj:CD (SD)	10	Testis weights were significantly decreased and seminiferous epithelial cells disappeared in the 2- and 4-week groups	Possible by a single administration	Nippon Boehringer Ingelheim Co.	Tsunenari et al.
Busulfan	p.o.	30, 100 and 300 mg/kg	Crj:CD (SD)	3	In the preliminary study, all animals died within 5 days and no abnormalities were observed at necropsy	Main study not conducted	Mitsubishi-Tokyo Pharmaceuticals	Not published
Compound C (platinum complex)	i.v.	10 and 20 for 2 weeks and 10 for 4 weeks and single dose of 40 and 80	Slc:SD	5	Administration period was shortened to 1 week due to marked weight loss in the 20 mg/kg group. Histopathologically, dilatation of the seminiferous tubules and degeneration/necrosis of the seminiferous epithelium were observed in the 2- and 4-week groups. Single administration also caused testicular lesions	Possible by a single administration or 2-week study	Chugai Pharmaceutical Co.	Misawa et al.

Cyclophosphamide	p.o.	5, 10, 20 and 40 for 2 weeks and 2.5, 5 and 10 for 4 weeks	Crj:CD (SD)	5	All animals in the 40 mg/kg group died during the treatment period. Reproductive organ weights did not change significantly in either 2- or 4-week groups. Quantitative analysis revealed testicular lesions in the 2- and 4-week groups	Possible by a 2-week study	Kowa Co.	Watanabe et al.
Cyclophosphamide	p.o.	Single dose of 100 mg/kg	Crj:CD (SD)	5	No effect on the reproductive organ weights. Histopathologically, decreased spermatogonia and preleptotene spermatocytes were observed	Possible by a single administration	Otsuka Pharmaceutical Co.	Matsumoto et al.
Etoposide	i.v.	5 and 10 mg/kg/week for 2 and 4 weeks	Crj:CD (SD)	6	Testis weights were decreased in the 4-week groups, but not in the 2-week groups. Histopathologically, decrease/loss of germ cells was observed in every treatment group. Histopathological changes in the testis were remarkable in the spermatogonia and spermatocytes	Possible by a 2-week study	Nikken Chemicals Co.	Kawaguchi et al.

(continued)

Table 9.1 (continued)

Compound	Route	Dose levels (mg/kg/day)	Animal strain	No. of animals/ group	Results	Evaluation	Company	Reference
Methyl methanesulfonate	p.o.	20 and 40 for 2 weeks and 20 for 4 weeks	Crj:CD (SD)	10	Decreased reproductive organ weights, decrease in germ cells, exfoliation of germ cells and vacuolar degeneration of Sertoli cells were observed in the 2- and 4-week groups	Possible by a 2-week study	Kissei Pharmaceutical Co.	Ozawa et al.
(4) *Cell division inhibitors*								
Carmofur (HCFU)	p.o.	200 for 2 and 4 weeks	Jcl:SD	5–15	In the 2-week group, the administration period was shortened to 11 days due to an excess mortality. Reproductive organ weights were decreased in both treatment groups. Histopathologically, vacuolar degeneration of Sertoli cells and exfoliation of round spermatids were observed in both treatment groups	Possible by a 2-week study	Mitsui Pharmaceutical Co.	Furukawa et al.
E7010 (a sulfonamide antitumor agent)	p.o.	50 and 75 for 2 weeks and single dose of 50	Slc:SD	5	Reproductive organ weights were decreased, and decrease/loss of seminiferous epithelial cells was observed in the 2-week group. The single dose caused decrease of round spermatids	Possible by a single administration or 2-week study	Eisai Co.	Hayakawa et al.

5-Fluorouracil	p.o.	20 and 30 for 2 weeks and 20 for 4 weeks	Slc:SD	5	Reproductive organ weights were decreased in all treatment groups. Histopathologically, degeneration of the seminiferous epithelium in the testis and degenerative spermatogenic cells in the epididymal ducts were observed in the 2-and 4-week groups	Possible by a 2-week study	Nippon Roche	Inomata et al.
(5) *Other drugs*								
Enoxacin	p.o.	3,000 for 2 weeks	Jcl:SD	5	Epididymis weights were decreased, and degeneration of spermatids and spermatocytes, retention of step 19 spermatids and multinucleated giant cell formation were observed	Possible by a 2-week study	Toyama Chemical Co.	Kizawa et al.
Nitrofurazone	p.o.	50 and 100 for 2 weeks and 50 for 4 weeks	Crj:CD (SD)	10	Reproductive organ weights were decreased, and seminiferous tubules in the testis were severely atrophied with a complete loss of spermatids and degeneration and desquamation of spermatocytes in all treated groups	Possible by a 2-week study	Yamanouchi Pharmaceutical Co.	Ito et al.

(continued)

Table 9.1 (continued)

Compound	Route	Dose levels (mg/kg/day)	Animal strain	No. of animals/ group	Results	Evaluation	Company	Reference
Pyrimethamine	p.o.	12.5, 25 and 50 for 2 weeks and 6.25, 12.5 and 25 for 4 weeks	Slc:SD	5	No changes in the weights of reproductive organs or in the number and activity of sperms in the epididymides were noted. Histopathologically, loss and degenerative changes of pachytene spermatocytes were observed in rats given 50 mg/kg for 2 weeks, but not in any group after 4 weeks	Possible by a 2-week study	Wyeth Lederle Japan	Murakami et al.
Theobromine	p.o.	250 and 500 for 2 and 4 weeks	Crj:CD (SD)	6	Reproductive organ weights were decreased, and histopathologically, degeneration/necrosis and desquamation of spermatids and spermatocytes were observed in the 2- and 4-week groups	Possible by a 2-week study	Dainippon Pharmaceutical Co.	Funabashi et al.
Theophylline	p.o.	300 and 600 for 2 weeks and 100 and 200 for 4 weeks	Crj:CD (SD) IGS	4–5	In the preliminary study, most of the animals died after single dose of 300 and 600 mg/kg, and no effects were observed on the testis weights and number of sperm after 4-week administration	Main study not conducted	Wyeth Lederle Japan	Not published

Theophylline	p.o.	150, 200 and 300 for 2 weeks and 150 for 4 weeks	Crj:CD (SD)	5	In the preliminary study, distention of seminiferous tubules was observed without dose dependence	Main study not conducted	Taiho Pharmaceutical Co.	Not published
(6) *General chemicals*								
Boric acid	p.o.	125, 250 and 500 for 2 and 4 weeks	Crj: Wistar	6	No effects on reproductive organ weights were observed. Sperm count and motile sperm rates were decreased in the 4-week study, but not in the 2-week study. Histopathologically, retention of step 19 spermatids was observed in the 2- and 4-week groups	Possible by a 2-week study	Kyorin Pharmaceutical Co.	Kudo et al.
Boric acid	p.o.	300 and 500 for 2 and 4 weeks	Jcl: Wistar	6	Decreased testis weights, exfoliation of round spermatids and retention of step 19 spermatids were observed in the 2- and 4-week groups	Possible by a 2-week study	Takeda Chemical Industries	Fukuda et al.

(continued)

Table 9.1 (continued)

Compound	Route	Dose levels (mg/kg/day)	Animal strain	No. of animals/ group	Results	Evaluation	Company	Reference
Dibromoacetic acid	p.o.	5, 50 and 250 for 2 weeks and 5 and 50 for 4 weeks	Crj:CD (SD)	6	Testis weights were not changed in any treated groups; however, the epididymal weights were significantly lower in the 250 mg/kg group. Histopathologically, atypical residual bodies and retention of step 19 spermatids were evident in the 250 mg/kg group. Retention of step 19 spermatids and increase in abnormal sperms were observed in both 50 mg/kg groups	Possible by a 2-week study	Mitsubishi-Tokyo Pharmaceuticals	Tsuchiya et al.
1,3-dinitrobenzene	p.o.	25, 50 and 75 for 2 weeks and 25 and 50 for 4 weeks	Crj:CD (SD)	4–5	Decreases in reproductive organ weights, decrease in the number of sperm, degeneration/necrosis of spermatogenic epithelium and giant cell formation were observed in the 2- and 4-week studies	Possible by a 2-week study	Taiho Pharmaceutical Co.	Irimura et al.
Ethylene glycol monomethyl ether	p.o.	100 and 200 for 2 weeks and 100 for 4 weeks	Crj:CD (SD)	6	Decreased reproductive organ weights, atrophy of seminiferous tubules and multinucleated giant cell formation were observed in the 2- and 4-week groups	Possible by a 2-week study	Asahi Chemical Industry	Watanabe et al.

From R. Bass, Proceedings of ICH 1, p. 283

Table 9.2 Comparison of NOAEL and ovarian toxicities observed in female fertility and general toxicity studies (Modified from Sanbuissho et al. 2009)

Chemicals		2-Week general toxicity study	4-Week general toxicity study	Reproductive study
Medroxyprogesterone acetate	NOAEL for ovarian toxicity	<0.4 mg/kg	<0.4 mg/kg	2.0 mg/kg
	Endpoints of ovarian or reproductive toxicity	Increase in atretic large follicles/ decreased currently/ previously corpora lutea	Increase in atretic large follicles/ decreased currently/ previously corpora lutea	Prolonged mean oestrous cycle before mating
	Ovarian toxicity or reprotoxicity was detected?	Yes	Yes	Yes
Mifepristone	NOAEL for ovarian toxicity	4 mg/kg	4 mg/kg	4 mg/kg
	Endpoints of ovarian or reproductive toxicity	Increase in luteinized cyst, increase in atresia of large follicle	Increase in luteinized cyst, Increase in atresia of large follicle	Persistent oestrus, shorter precoital interval, completely infertile animals
	Ovarian toxicity or reprotoxicity was detected?	Yes	Yes	Yes
Tamoxifen	NOAEL for ovarian toxicity	0.03 mg/kg	0.005 mg/kg	<0.005 mg/kg
	Endpoints of ovarian or reproductive toxicity	Hyperplasia of the interstitial cell, increase in atretic follicle, large sized, absence of currently corpus luteum	Hyperplasia of the interstitial cell, Increase in atretic follicle, large sized, Absence of currently corpus luteum	Increase in anestrus females, preimplantation loss
	Ovarian toxicity or reprotoxicity was detected?	Yes	Yes	Yes
4-Vinylcyclohexene diepoxide	NOAEL for ovarian toxicity	20 mg/kg	5 mg/kg	20 mg/kg
	Endpoints of ovarian or reproductive toxicity	Decrease in small follicle	Decrease in small follicle	Increase in preimplantation loss
	Ovarian toxicity or reprotoxicity was detected?	Yes	Yes	Yes

(continued)

Table 9.2 (continued)

Chemicals		2-Week general toxicity study	4-Week general toxicity study	Reproductive study
Busulfan	NOAEL for ovarian toxicity	2.5 mg/kg	0.5 mg/kg	0.5 mg/kg
	Endpoints of ovarian or reproductive toxicity	No noteworthy findings	Decrease of small follicles	Increase in preimplantation loss
	Ovarian toxicity or reprotoxicity was detected?	No	Yes	Yes
Cisplatin	NOAEL for ovarian toxicity	0.5 mg/kg	0.125 mg/kg	0.25 mg/kg
	Endpoints of ovarian or reproductive toxicity	Increase in atresia of medium and large follicle, decrease in large follicle and newly formed corpus luteum	Increase in atresia of medium and large follicle, Decrease in large follicle and newly formed corpus luteum	Increase in preimplantation loss
	Ovarian toxicity or reprotoxicity was detected?	Yes	Yes	Yes
Cyclophosphamide	NOAEL for ovarian toxicity	20 mg/kg	10 mg/kg	<5 mg/kg
	Endpoints of ovarian or reproductive toxicity	Not noteworthy findings	Atrophy of corpus luteum	Increased preimplantation loss
	Ovarian toxicity or reprotoxicity was detected?	No	Yes	Yes
Anastrozole	NOAEL for ovarian toxicity	1 mg/kg	0.1 mg/kg	0.01 mg/kg
	Endpoints of ovarian or reproductive toxicity	Increase in large-sized atretic follicle, cystic follicle	Increase in large-sized atretic follicle	Irregular oestrous cycle, increased preimplantation loss
	Ovarian toxicity or reprotoxicity was detected?	Yes	Yes	Yes
Di(2-ethylhexyl)adipate	NOAEL for ovarian toxicity	200 mg/kg	200 mg/kg	200 mg/kg
	Endpoints of ovarian or reproductive toxicity	Increase in large-sized atretic follicle	Increase in large-sized atretic follicle	Increase of oestrous cycle and preimplantation loss
	Ovarian toxicity or reprotoxicity was detected?	Yes	Yes	Yes

Di(2-ethylhexyl) phthalate	NOAEL for ovarian toxicity	1,000 mg/kg	300 mg/kg	1,000 mg/kg
	Endpoints of ovarian or reproductive toxicity	Increase in large-sized atretic follicles	Increase in large-sized atretic follicles. Decrease in newly formed corpora lutea	Prolonged oestrous cycle
	Ovarian toxicity or reprotoxicity was detected?	Yes	Yes	Yes
Ethylene glycol monomethyl ether	NOAEL for ovarian toxicity	30 mg/kg	30 mg/kg	<30 mg/kg
	Endpoints of ovarian or reproductive toxicity	Hypertrophy of corpora lutea, Increased large atretic follicles	Hypertrophy of corpora lutea, Increased large atretic follicles	Irregular oestrous cycle
	Ovarian toxicity or reprotoxicity was detected?	Yes	Yes	Yes
Indomethacin	NOAEL for ovarian toxicity	1.3 mg/kg	1.3 mg/kg	1.3 mg/kg
	Endpoints of ovarian or reproductive toxicity	Unruptured follicle with luteinizing	Unruptured follicle with luteinizing, Increase in atretic follicle	Not evaluated. 4.0 mg/kg: death 8/10 1.3 mg/kg: no change
	Ovarian toxicity or reprotoxicity was detected?	Yes	Yes	–
Compound X	NOAEL for ovarian toxicity	<4 mg/kg	<4 mg/kg	4 mg/kg
	Endpoints of ovarian or reproductive toxicity	Increase in atresia of large follicle	Increase in atresia of large follicle	Decrease in corpora lutea, implantations and live embryo
	Ovarian toxicity or reprotoxicity was detected?	Yes	Yes	Yes
Atrazine	NOAEL for ovarian toxicity	30 mg/kg	30 mg/kg	3 mg/kg
	Endpoints of ovarian or reproductive toxicity	Increase in large atretic follicle, loss of currently formed corpus luteum, decrease in previously formed, swelling of previously	Increase in large-sized atretic follicle swelling of previous formed corpus luteum	Increase of irregular oestrous cycle
	Ovarian toxicity or reprotoxicity was detected?	Yes	Yes	Yes

(continued)

Table 9.2 (continued)

Chemicals		2-Week general toxicity study	4-Week general toxicity study	Reproductive study
Bromocriptine	NOAEL for ovarian toxicity	0.08 mg/kg	0.08 mg/kg	0.08 mg/kg
	Endpoints of ovarian or reproductive toxicity	Increase in corpora lutea	Increase in absolute ovary weight, Increase in corpora lutea	Decrease in pregnant animals
	Ovarian toxicity or reprotoxicity was detected?	Yes	Yes	Yes
Chlorpromazine hydrochloride	NOAEL for ovarian toxicity	3 mg/kg	3 mg/kg	3 mg/kg
	Endpoints of ovarian or reproductive toxicity	Increase of large-sized atretic follicle	Increase of large-sized atretic follicle	Irregular oestrous cycle
	Ovarian toxicity or reprotoxicity was detected?	Yes	Yes	Yes
Sulpiride	NOAEL for ovarian toxicity	<1 mg/kg	<1 mg/kg	1 mg/kg
	Endpoints of ovarian or reproductive toxicity	Follicular cyst, increase of large-sized atretic follicle	Follicular cyst, Increase of large-sized atretic follicle	Irregular oestrous cycle
	Ovarian toxicity or reprotoxicity was detected?	Yes	Yes	Yes

- *Flexibility of the approach presented in the ICH reproductive toxicity guideline*
  Whereas normally the design for using *three distinct segments for testing* will be used and will suffice for screening purposes, it remains a case-by-case approach, which needs to be balanced against all other possible designs (1) treatment period(s), from single days used for treatment to treatment over several generations (in one go). (2) Choice of doses ranging from totally non-toxic to the parental and maternal generations *and* the offspring to seriously intoxicating either/or both parental and offspring generations may become necessary, as well as increasing the numbers of doses above three. (3) In-depth investigations of both kinetic and dynamic behaviour of the test substance in adult and developing organisms and species may become necessary. (4)Imagination in finding, refining and validating experimental endpoints may be of utmost importance. It is certainly impossible to address all of these for each and every medicinal product. Scientific inquisitiveness and expertise will help guide to doing the best possible approach to understanding non-/toxicity to reproduction.
- *Highlights from the first ICH guideline*
  "The combination of studies selected should allow exposure of mature adults and all stages of development from conception to sexual maturity. To allow detection of immediate and latent effects of exposure, observations should be continued through one complete life cycle, i.e. from conception in one generation through conception in the next generation" (from the guideline). The stages described allow for setting up segmental studies addressing one or several of the following: premating to conception, conception to implantation, implantation to closure of the hard palate, closure of the hard palate to the end of pregnancy, birth to weaning and weaning to sexual maturity.

  Following consideration of all available data (pharmacological, kinetic and toxicological—for the test and related substance) *the most probable option* will usually be adequate for testing. This can be equated to a combination of studies for effects on:

  Segment 1: Fertility and early embryonic development
  Segment 2: Embryo–fetal development
  Segment 3: Pre- and postnatal development—including maternal function.

  See Figs. 9.3, 9.4 and 9.5.

  This was accepted at ICH 2 in 1993 and has been applied successfully ever since.

  Whereas male fertility still needed to be investigated in a separate set of animals, scientific impetus was to include this study into the repeat-dose study (and to come back to special male fertility investigations only upon suspicion raised from such repeat-dose studies).
- *From the first harmonised tripartite guideline to complete harmonization of reproductive toxicity testing*.

Continuation of Takayama's arguments:

4. Anything beyond a core package composed of EC segment 1 and Japanese segments 2 and 3 studies would be a waste of money, time and animals.
5. Whereas normal requirements are appropriate for screening, any problem arising would need to be addressed in more detail anyway.

Whereas the first ICH guideline on reproductive toxicity testing still lacked complete harmonization concerning routine approaches and testing, namely, in the area of segment 1, ongoing evaluation of study data in Japan allowed for agreement later in this area: under normal conditions, it has now become possible to include male fertility testing into repeat-dose studies (and the conditions for achieving this option are laid out). When data became available to the Japanese colleagues, the two tripartite guidelines on reproductive toxicity testing (the second one dealing only fertility) were combined into one guideline.

Other areas of concern address testing approaches using alternatives to animals (rodents like mice and rats, non-rodents like rabbits and other non-human species like pigs and non-human primates): *non-whole animal systems* (using organ-, tissue- and cell-cultures), like the mouse Embryonic Stem Cell Test (EST). Whereas various processes of early human reproduction can be performed or mimicked ex vivo, the complex sequence of steps representing reproduction cannot.

9.5.7 The End

With the inclusion of the addendum on male fertility testing into the main guideline body, the harmonization process was finished for the area of reproductive toxicity testing in 2000. Even though the scientific positions already in 1991 allowed for achieving harmonization, it took 2 years for the harmonization of testing with the caveat about male fertility, which was finally resolved 7 years later. The outlook into the ICH future was in two directions. One was to investigate the opportunities to apply in vitro models and incorporate them into routine testing. The current state of the art is described in the report from a recent workshop (van der Laan et al. 2012); this report shows the scientific strive towards achieving better information using less animals—and the difficulties experienced. The second one was to expand harmonization from agreeing conditions for marketing authorisation to initiation and continuation of clinical trials in healthy human volunteers, patients including women of childbearing potential and fertile men. This required the inclusion not only of scientific reasoning but also of political impact and *equal rights* for women and men to be included in early clinical trials to benefit from new treatments as early as possible. This required not only reshuffling of the previous ICH S5(R2) reproductive toxicity testing guideline but also introduction of compromises; the current state is described by Sjöberg and David (2013).

References

Bass R (1991) Proposal for the development of a uniform tripartite guideline. In: D'Arcy and Harron (eds) Proceedings of the first international conference on harmonization, Brussels, pp 279–288 (discussion: 289–293)

Bass R (1994) Report on progress since ICH 1: Safety (2), Development of the reproductive toxicology guidelines to become one of the first to complete the ICH process. In: D'Arcy and Harron (eds) Proceedings of the second international conference on harmonization, Brussels, pp 33–38

Bass R, Ulbrich B (1991) Draft guideline on detection of toxicity to reproduction for medicinal products – Draft No. 12. Adverse Drug React Toxicol Rev 9(3):127–141

Bass R et al (1991) Draft guideline on detection of toxicity to reproduction for medicinal products – Draft No. 17. Adverse Drug React Toxicol Rev 10(3):143–154

Christian MS (1992) Harmonization of reproductive guidelines: perspective from the International Federation of Teratology Societies. J Am Coll Toxicol 11(3):299–302

Christian MS (2001) Test methods for assessing female reproductive and developmental toxicology. In: Wallace Hayes A (ed) Principles and methods of toxicology, 4th edn. Taylor & Francis, Philadelphia, PA, pp 1301–1381

Council Directive 65/65/EEC (of 26 January 1965) On the approximation of provisions laid down by law, regulation or administrative action relating to medicinal products. OJ L 22:369

Council Directive 75/318/EEC (of 20 May 1975) On the approximation of the laws of the Member States relating to analytical, pharmaco-toxicological and clinical standards and protocols in respect of the testing of proprietary medicinal products. OJ L 147:1–12

Diener RM (1991) Timing and utility of behavioural studies in developmental toxicology. In: D'Arcy PF, Harrison (eds) Proceedings of the first international conference on harmonization, Brussels, pp 273–279

European Union (EU)-Legislation: EUDRALEX: Volume 3: Scientific guidelines for medicinal products for human use: see now EMA Web site

ICH S5(R2) (2000) Safety guidelines: detection of toxicity to reproduction for medicinal products and toxicity to male fertility. In: ICH Web site (IFPMA)

ICH S5a (1993) Harmonised tripartite guideline on detection of toxicity to reproduction for medicinal products

ICH S5a(R1) (1995) Toxicity to male fertility, an addendum to the ICH tripartite guideline on detection of toxicity to reproduction for medicinal products

Jödicke B, Neubert D (2004) Reproduktion und Entwicklung. In: Marquardt, Schäfer (eds) Lehrbuch der Toxikologie, chap 21, 2nd edn. Wissenschaftliche Verlagsgesellschaft, Stuttgart, pp 419–544

Lumley CE (1991) Proposal for international guidelines for reproductive ad developmental toxicity testing for pharmaceuticals. Adverse Drug React Toxicol Rev 10(3):143–153

M3 (1997) Non-clinical safety studies for the conduct of human clinical trials for pharmaceuticals

M3(R2) (2009) Guidance on nonclinical safety studies for the conduct of human clinical trials and marketing authorization for pharmaceuticals. In: ICH Web site (IFPMA)

MHLW (1989) Notification: Report of review of the guideline for the detection of influence of pharmaceuticals on reproduction. Yakushin No 118 (1989.9.11)

OECD (1981) Guidelines for testing of chemicals. Sect 4: Health effects. Guideline No. 414: Teratogenicity. OECD, Paris

OECD (1983a) Guidelines for testing of chemicals. Sect 4: Health effects. Guideline No. 415: One generation reproduction toxicity study. OECD, Paris

OECD (1983b) Guidelines for testing of chemicals. Sect 4: Health effects. Guideline No. 416: Two generation reproduction toxicity study. OECD, Paris

Omori Y (1991) Principles and guidelines—a review of recommendations (on detection of toxicity) in the three regions. In: D'Arcy PF, Harrison (eds) Proceedings of the first international conference on harmonization, Brussels, pp 256–266

Peters PWJ, Garbis-Berkvens JM (1996) Methods in reproductive toxicology. In: Niesink RJM et al (eds) Toxicology: principles and applications. CRC, Boca Raton, FL, pp 949–974

Sakai T et al (2000) Collaborative work to evaluate toxicity on male reproductive organs by 2-week repeated dose toxicity studies in rats. J Toxicol Sci 25(Spl Issue):1–21

Sanbuissho A et al (2009) Collaborative work on evaluation of ovarian toxicity by repeated-dose and fertility studies in female rats. J Toxicol Sci 34(Spl Issue):SP1–SP22

Sjöberg P, David RJ (2013) Non-clinical safety studies for the conduct of human clinical trials for pharmaceuticals—ICH M3 and M3(R2). In: van der Laan JW, DeGeorge JJ (eds) Global approach in safety testing. AAPS advances in the pharmaceutical sciences series 5. Springer, New York

Sullivan FM, Watkins WJ, van der Venne MTh (eds) (1993) The toxicology of chemicals. Series two: Reproductive toxicology, vol 1: Summary reviews of the scientific evidence. EUR 12029 EN 14991

Takayama S (1991) Proposal or mutual acceptance of studies. In: D'Arcy PF, Harrison (eds) Proceedings of the first international conference on harmonization, Brussels, pp 266–269

Takayama S et al (1995) Studies on the optimal treatment period and parameters for detection of male fertility disorder in rats-introductory summary. J Toxicol Sci 20:73–82

Van Cauteren H (1991) Panel presentation. In: D'Arcy PF, Harrison (eds) Proceedings of the first international conference on harmonization, Brussels, p 289

van der Laan JW et al (2012) Testing strategies for embryo-fetal toxicity of human pharmaceuticals. Animal models vs. in vitro approaches—a workshop report. Regul Toxicol Pharmacol 63:115–123

Weissinger J (1991) Commentary on proposal for mutual acceptance and proposed alternative approaches. In: D'Arcy PF, Harrison (eds) Proceedings of the first international conference on harmonization, Brussels, pp 269–273

Chapter 10
ICH S6 Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals

Joy Cavagnaro and Jennifer Sims

Abstract Since its publication in 1997, the ICH guideline on the Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals ("ICH S6") has fostered consistency while maintaining the necessary flexibility for testing within and across a variety of product classes. Successful implementation of a product-specific science-based "case-by-case" approach however has required individuals with a broad knowledge of toxicological processes and the ability to integrate data from molecular biology, pharmacology, physiology, pharmacokinetics, and pathology. Importantly, the "case-by-case" approach only works if there is an understanding of the science and an acceptance by both regulators and industry that the interpretation of the data has to reflect best scientific practice and that no study in experimental animals can predict with certainty the outcome in humans. As such, a greater dialogue between industry and regulatory authorities has been needed early and in some cases throughout development to ensure that the decision on how a product should be tested not only meets the stringencies of the regulatory authorities but is also designed to improve the predictive value for extrapolating to humans. This dialogue between industry and regulatory authorities continued to the ICH Expert Working Group charged with formulating the addendum to ICH S6(ICH S6R(1) finalized at step 4 in June 2011), guidance based on the accumulated and collective experience of the safety assessment of biotechnology-derived pharmaceuticals in the 14 years since the finalization of ICH S6.

J. Cavagnaro (✉)
Access BIO, Boyce, VA, USA
e-mail: jcavagnaro@accessbio.com

J. Sims
Integrated Biologix GmbH, Basel, Switzerland

J.W. van der Laan and J.J. DeGeorge (eds.), *Global Approach in Safety Testing*, AAPS Advances in the Pharmaceutical Sciences Series 5, DOI 10.1007/978-1-4614-5950-7_10,

10.1 Introduction

10.1.1 Historical Perspective

In the early 1980s, neither industry toxicologists nor regulatory scientists were sure of what constituted an appropriate toxicological assessment program for biopharmaceuticals. There were even some who believed that natural proteins were inherently safe thus the toxicity should be minimal or not relevant. However, in 1986, the biotechnology working party was established in Europe to focus on specific issues related to the development of biotechnology-derived pharmaceuticals. In July of that same year, a satellite symposium to the IV International Congress of Toxicology was held at the Keio Plaza Hotel, Tokyo, Japan. Attendees included government regulatory scientists, university scientists, and industrial scientists and research managers, all with an interest in the development of new biotechnology-derived products (Giss 1987; Dayan 1987; Galbraith 1987; Finkle 1987; Zbinden 1987).

10.1.2 Proposal for a Specific Guidance for Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals

Five years later, at the first ICH meeting in Brussels, Belgium, in 1991, it was questioned whether differing attitudes among the various regions towards development of biotechnology-derived pharmaceuticals were considered significant enough to actually justify a session. A session was held, and a "rational science-based approach" was acknowledged as critical to the successful and expeditious development of new and novel products (Hayakawa 1992; Cavagnaro 1992a; Hohbach 1992). One of the issues addressed in the workshop was whether common standards and attitudes that were evolving could be maintained without the issuance of formal guidance. The recommendations from the workshop were that in the short term, regulatory authorities should maintain a flexible approach to requirements for preclinical testing on a case-by-case basis, and the work should be initiated to prepare internationally accepted principles for the safety evaluation of drugs produced using biotechnology (Kikuchi 1992). Importantly, even in the early 1990s, it was recognized that the value of case-by-case for globalizing markets depended fully on a common understanding of all partners involved. If this was not achieved, there would be a continuous risk for inequality of advice on the requirements and standards from one country to another.

Supporting publications were also emerging questioning the relevance of the traditional pharmaceutical paradigm for the preclinical safety evaluation of biopharmaceuticals (Zbinden 1990, 1991; Bass et al. 1992; Hayes and Cavagnaro 1992; Cavagnaro 1992b; Claude 1992; Terrell and Green 1994; Dayan 1995; Thomas 1995; Henck et al. 1996). During this time period, there were both increases in the number of biopharmaceuticals under development and a rapidly increasing number of small companies coming into the field. At the second ICH

Table 10.1 Members of the ICH S6 Expert Working Group

Europe	Japan	United States
Dr. Jennifer Sims (EU)	Dr. Tohru Inoue (MHW)	Dr. James Green (PhRMA)
Prof. Giuseppe Vicari (EU)	Dr. Mashiro Nakadate (MHW)	Dr. Joy Cavagnaro (FDA) (Rapporteur)
Dr. Jorgen Carstensen (EFPIA)	Dr. Eliji Makai (JPMA)	
Dr. Wolfgang Neumann (EFPIA)	Dr. Mutsufumi Kawai (JPMA)	

meeting in Orlando, Florida, in 1993, biotechnology issues mainly focused on product quality issues although interest was increasing with rumblings for a more formal guidance for preclinical assessment of biotechnology-derived pharmaceuticals. Soon after this meeting, an ICH Expert Working Group (EWG) was established, and a concept paper was proposed by the FDA. A pre-step 2 document was released at the third ICH meeting in Yokohama, Japan, in 1995. A few years later in February of 1997, the 13th CMR International Workshop provided an opportunity for international experts to discuss experiences and difficulties encountered in designing scientifically based preclinical safety evaluation programs for biopharmaceuticals. This 2-day meeting brought together toxicologists and clinicians, from 32 pharmaceutical and biotechnology companies and regulators and regulatory advisors from the European Agency for the Evaluation of Medicines (EMEA, now European Medicines Agency, EMA) and 9 countries: Denmark, France, Germany, Italy, Japan, the Netherlands, Sweden, the UK, and the USA (Griffith and Lumley 1998). Recommendations arising from the CMR Workshop were taken into consideration by the expert working group for the final drafting of ICH S6 guideline, and agreement was reached at ICH 4 in Brussels in July 1997 (ICH S6 Preclinical Safety Evaluation of Biotechnology-derived Pharmaceuticals 1997) (Table 10.1).

10.1.3 Implementation of ICH S6

Over the ensuing decade, the numbers, types, complexities, and indications for "biotech products" grew. Many of these novel products were successfully approved for market. Publications provided insight into experiences with the case-by-case approach strategies (Serabian and Pilaro 1999; Sims 2001; Ryan and Terrell 2002; Cavagnaro 2002; Brennan et al. 2004; Buckley et al. 2008). However, the explosion in new constructs and novel formats was also complicated with the arrival of second-generation products in the form of "biosimilars" and "biobetters" of the first-generation products approved for use in the 1990s. In parallel to the industry evolution, some key regulatory agencies also underwent reorganization, and there were also changes in industry access to regulatory authorities for informal and formal dialogue. This industry-regulatory evolution resulted in a combined industry-regulatory "creep" in terms of preclinical development programs to support biopharmaceuticals. A trend started to

emerge for an increasing number of questionable studies and the application of ICH guidance documents to biopharmaceuticals where biopharmaceuticals were specifically excluded in the scope of such guidance. There was also a concern for potential increases in regional guidance to aid in interpretation of ICH S6 (Nakazawa et al. 2004).

10.1.4 Rationale for Updating ICH S6

While there were reservations by some that updating ICH S6 could result in formalizing the emerging increase in studies, the perception of a considerable drift in the interpretation and application of the original intent of the ICH S6 guidance led to a series of regional industry-regulatory scientific meetings in June of 2007 to discuss specific topics identified as issues when applying the S6 guidance. The conclusions of these meetings were the need to evaluate the state of the art of safety testing of biopharmaceuticals. During this time under the auspices of BioSafe, a series of white papers were published on a series of topics (e.g., tissue cross-reactivity, species selection, immunogenicity, reproductive toxicity, carcinogenicity) and a review of scientific state-of-the-art best practice was published in Preclinical Safety Evaluation of Biopharmaceuticals: a science-based approach to facilitating clinical trials (ICH S6R). These publications would provide the necessary background for deliberations of the new ICH S6 EWG (Table 10.2).

10.1.5 Addendum to ICH S6: ICH S6(R1)

In June 2008, the ICH Steering Committee endorsed a concept paper on the proposal to establish an EWG to write an addendum to the ICH S6 guidance—the ICH S6R(1) addendum. The concept paper stated that there was a need for a clarification (and sometime amplification) of ICH S6 since substantial experience and new information has been gained since *step 4* (1997). The preclinical safety experts involved in ICH in S2/S9/M3 agreed that the flexible and case-by-case approach described in the original guidance is still valid and must be preserved. Based on the outcome of these discussions, it was agreed that the following topics would be addressed to facilitate the understanding and harmonized application of the guidance provided in S6:

- Species selection
  - How to justify the choice of a species
  - Clarify the role of tissue cross-reactivity
  - When to use a second species
  - Use of alternative models such as transgenics and homologous products

Table 10.2 Key papers outlining experiences and proposed best practices for preclinical assessment of biopharmaceuticals

Duration of chronic toxicity studies for biotechnology-derived pharmaceuticals: is 6 months still appropriate? (Clarke et al. 2008)[a]
Scientific review and recommendations on preclinical cardiovascular safety evaluation of biologics (Vargas et al. 2008)
Alternative strategies for toxicity testing of species-specific biopharmaceuticals (Bussiere et al. 2009)[a]
Consideration in assessing the developmental and reproductive toxicity potential of biopharmaceuticals (Martin et al. 2009)[a]
Practical approaches to dose selection for first-in-human clinical trials with novel biopharmaceuticals (Tibbitts et al. 2010)[a]
Use of tissue cross-reactivity studies in the development of antibody-based biopharmaceuticals: history, experience, methodology, and future directions (Leach et al. 2010)[a]
Carcinogenicity assessments of biotechnology-derived pharmaceuticals: a review of approved molecules and best practice recommendations (Vahle et al. 2010)[a]
Developmental toxicity testing of biopharmaceuticals in nonhuman primates: previous experience and future directions (Martin and Weinbauer 2010)
Preclinical safety evaluations supporting pediatric drug development with biopharmaceuticals: strategy, challenges, current practices (Cavagnaro 2008a)[a]

[a]Developed by ad hoc committees of the Biotechnology Industry Organization's (BIO) Preclinical Safety (BioSafe) Committee. BioSafe is a committee within the BIO Health Section. BioSafe is composed of BIO members working to serve as a resource for BIO members and BIO staff by identifying key scientific and regulatory issues and developments related to the preclinical safety evaluation of biopharmaceutical products and recommending appropriate science-based responses

- Study design
  - Scientific justification of duration of chronic toxicity study
  - High dose selection
  - Utility and length of recovery
- Reproductive/developmental toxicity
  - Justification of species selection including the use of rodents versus non-rodents and use of alternative models such as transgenics and homologous products
  - Considerations when using primates: use of combined study designs and timing of these studies, how to get data on fertility, impact of placental transfer, and how to get data from the F1 generation
- Carcinogenicity
  - Justification for the approach to address carcinogenic risk
  - Application of in vivo models: length of studies, use of proliferation indices, and use of homologous products
- Immunogenicity
  - Extent of characterization
  - Impact of neutralizing versus non-neutralizing

Table 10.3 Members of the ICH S6(R1) Expert Working Group

Europe	Japan	United States
Dr. Jan-Willem van der Laan (EU) (rapporteur from step 2)	Dr. Yoko Hirabayashi (NIHS)	Dr. Ann Pilaro (FDA)
Dr. Beatriz Silva Lima (EU)	Dr. Kazushige Makai (PMDA)	Ms. Mercedes Serabian (FDA)
Dr. Jennifer Sims (EFPIA) (rapporteur to step 2)	Dr. Matusmoto Mineo (PMDA)	Dr. Abigail Jacobs (FDA)
Dr. Maggie Dempster (EFPIA)	Dr. Takahiro Nakazawa (JPMA)	Dr. David Jacobson-Kram (FDA)
	Dr. Atsushi Sanbuissho (JPMA)	Dr. Ruth Lightfoot-Dunn (PhRMA)
	Dr. Kazuichi Nakamura (JPMA)	Dr. Helen Haggerty (PhRMA)

- Role of PD markers
- Assessment of recovery groups

ICH S6R(1) was finalized under *step 4* in June 2011. The harmonized addendum provides further complementary guidance to the S6 guidance and helps to define current recommendations and, hopefully, should reduce the likelihood that substantial differences will exist among regions. The addendum ICH S6R(1) is integrated as part II in the core S6 guideline (ICH S6R) (Table 10.3).

10.2 Definition of Biotechnology-Derived Pharmaceutical

The initial ICH S6 guidance was intended to recommend a basic framework for preclinical safety evaluation of biopharmaceuticals. Biotechnology-derived pharmaceuticals were defined as products derived from characterized cells including bacteria, yeast, insect, plant, and mammalian cells. The active substances include cytokines, growth factors, fusion proteins, toxin conjugates, enzymes, clotting factors, thrombolytics, soluble receptors, hormones, and monoclonal antibodies (ICH S6 Preclinical Safety Evaluation of Biotechnology-derived Pharmaceuticals 1997). Importantly it was recognized that with each product class, there may also be variations. For example, over the years, monoclonal antibody products would evolve to include murine, chimeric, humanized, and fully humanized as well as "antibody-like" molecules and antibody derivatives. Products would span monospecific, bispecific, or trispecific variants; naked or conjugated; antagonist, agonist, or catalytic; targeting an endogenous epitope or a foreign epitope; with unique species specificity or with broad specificity; with no target or off-target binding on any "normal" animal species; or with specific binding to an epitope which is only upregulated in the disease state.

It was acknowledged that the principles outlined in the guidance may also be applicable to recombinant DNA protein vaccines, chemically synthesized peptides, plasma-derived products, endogenous proteins extracted from human tissue, and oligonucleotide-based drugs (ICH S6 Preclinical Safety Evaluation of Biotechnology-derived Pharmaceuticals 1997).

10.3 Key Differences Between Biopharmaceuticals and Pharmaceuticals

Biopharmaceuticals and pharmaceuticals can be viewed as a product continuum based on size and complexity in molecular structure. However as products have evolved, there has been a blurring of product attributes. Small molecules have become larger as the result of alternative scaffolding technologies, e.g., protein conjugates and fusion proteins in order to improve exposure characteristics and dosing regimens. Large molecules have become smaller, e.g., antibody fragments and protein mimetics in order to improve distribution and decrease potential immunogenicity (Cavagnaro 2010). Novel delivery technologies are also enabling alternative routes of delivery for biopharmaceuticals, e.g., by the oral and inhalation routes. Some products such as oligonucleotide-based drugs (ONs) may have combined product attributes. For example, ONs are synthetically derived but have complex chemical profiles and are catabolized in ways similar to those followed by certain biopharmaceuticals. Although toxicity assessments are designed to address hybridization-independent effects, some ONs can also exhibit species specificity where analogous sequences may be needed to assess hybridization-dependent effects, i.e., toxicity related to exaggerated pharmacology. Thus, specific considerations are based upon product class and product attributes that influence program design. Table 10.4 provides a general comparison of product attributes across product classes. While there will be exceptions, the general distinctions provide the rationale for the different approaches to preclinical safety evaluation.

10.4 Key Considerations of ICH S6

A seminal principle of ICH S6 is that safety evaluation programs should include relevant species demonstrating pharmacological activity. Thus, a key challenge in the preclinical evaluation of biopharmaceuticals is species specificity. Unlike pharmaceuticals, one cannot assume that a molecule will be active in two species, e.g., rodent (rat or mouse) and non-rodent (rabbit, dog, nonhuman primate) traditionally used for toxicity testing. An even greater challenge is when a product is uniquely species specific, i.e., it is only pharmacologically active in humans. Determining biological activity is based on an understanding of in vitro receptor occupancy, affinity, and distribution and in vitro and in vivo pharmacological effects. Importantly, toxicity studies in nonrelevant species were discouraged (ICH S6 Preclinical Safety Evaluation of Biotechnology-derived Pharmaceuticals 1997).

In general, 6-month duration for chronic dose studies was considered sufficient. However, it was acknowledged that specific considerations may require a longer duration study in some cases and shorter duration may also be acceptable in some cases. For example, formation of neutralizing antibodies could limit utility of longer-term dosing if there is significant impact on exposure.

Table 10.4 Comparative product attributes across product classes

Product attribute	Pharmaceutical	Oligonucleotide	Biopharmaceutical
Manufacture	Chemical synthesis	Chemical synthesis	Biological synthesis (cell culture, transgenic plants, transgenic animals)
Composition	Commonly organic chemical	Modified nucleic acids	Protein, carbohydrate
Size	Generally <1 kDa	Generally between 1 and 10 kDa	Generally >30 kDa up to 800 kDa
Purity	Homogeneous, single entity high chemical purity (except racemic mixtures)	Mostly single entity with product-related impurities	Heterogenous mixture(microheterogeneity, aggregates)
Potency	Not determined	Not determined	Required. Generally in vivo. For some product, in vitro is acceptable
Route of administration	Oral, topical, inhalation	Parenteral, inhalation	Parenteral or targeted (e.g., intracardiac, intrathecal, and intra-articular)
Pharmacokinetics	Half-life usually minutes to hours	Half-life usually hours to days	Half-life usually days to weeks. PK can be impacted by target leading to target-mediated drug disposition (TMDD)
ADME	Rapid entry through blood capillaries, distribution to many organs/tissues, metabolized to active and non-active metabolites	Catabolized to nucleotides and other metabolites depending on modifications, selected tissue distribution	Distribution limited based on size to plasma and/or extracellular fluid, degraded (catabolism) to endogenous amino acids
Drug–drug interactions	Influenced by metabolism enzyme induction or inhibition can be significant, transporters	Not metabolized by P450	Interactions related to additive or synergistic pharmacological activity not P450-related metabolism
Dose response	Linear establishment of maximum tolerated dose (MTD)	Linear establishment of maximum tolerated dose (MTD)	Can be bell-shaped establishment of optimum biological dose (OBD), MFD
Targets	Intracellular or ligand–receptor	Intracellular and extracellular	Cell–matrix, cell–cell, or ligand–receptor
Species specificity	Species independent, preclinical assessments for general toxicity generally performed in one rodent (generally rat) and one non-rodent (generally dog)	Hybridization specific and nonspecific assessments Use of analogous/homologous molecules	Species specific. Nonhuman primates often the only relevant species Use of homologous/surrogate molecules

Toxicological effects	Unpredictable, related to MOA, chemistry, metabolites	Mostly related to chemistry	Often predictable based on known mechanism of action "exaggerated pharmacology." Animal models of disease are often used to asses safety in addition to activity
Immunogenicity (not including allergic or hypersensitivity reactions)	Never	Rare	Common. May effect exposure and exposure to bioactive drug (neutralizing antibodies)
Dosing interval for chronic use	Daily	Intermittent	Intermittent
Dose formulation	Complex	Complex, simple	Simple

Adapted Cavagnaro (1992a, b, 2002, 2008a, b, 2010)

During the implementation of ICH S6, there was a misconception that only one species was expected for assessing general toxicity of biopharmaceuticals. However, the language in ICH S6 explicitly stated "that safety evaluation programs should normally include two relevant species but, in certain justifiable cases, one relevant species may suffice (e.g., when only one relevant species can be identified or where the biological activity of the biopharmaceutical is well understood)." Importantly, the guidance intentionally did not specify use of the "most relevant" in order to avoid the routine consideration of use of higher primate species (e.g., greatest homology of a protein or a receptor with chimpanzees or baboons). There was also a growing confusion on how to define a relevant species.

10.5 Key Developments in Study Design Since ICH S6

The scientific discussions and guidance in the ICH S6 addendum ICH S6R(1) drafted by the ICH Expert Working Group was based on the accumulated experience of industry and regulators over the 14 years since ICH was finalized in 1997. A number of literature reviews on various aspects of the preclinical safety evaluation of biotechnology-derived products (see Table 10.2) were considered as well as anonymized case studies from the regulatory databases and the impact of the 2006 "Tegenero" incident in the United Kingdom.

10.5.1 Number of Species

The number of species required for safety assessment became a growing industry concern. In large part because there were requests by regulatory authorities for rodent studies with homologous products or rodent toxicology studies where the species was not a pharmacologically relevant species to satisfy the requirement for two species as standard for pharmaceuticals. The addendum therefore clarified that if there are two pharmacologically relevant species for the clinical candidate (one rodent and one non-rodent), then both species should be used for short-term general toxicology studies. The use of one species for all general toxicity studies is justified when the clinical candidate is pharmacologically active in only one species, generally the nonhuman primate. However, in such cases, where the only relevant species is the nonhuman primate, studies in a second species with a homologous product are not considered to add further value for risk assessment and are not recommended (ICH S6R).

If two relevant species exist, then short-term repeat dose toxicity studies are recommended. However, if the target organ profile is similar across species and/or similar class, effects are observed, and the dose selected in the clinical trials appears acceptable, then chronic toxicity studies in a single species may be justifiable.

10.5.2 Selection of Relevant Species

Clarification is provided in the addendum on the scientific data required to support the selection of a relevant species for safety assessment. This includes an evaluation of cross-species sequence homology, in vitro target binding and functional activity data, and in vivo pharmacodynamics markers such as evidence for target engagement, modulation of a known biological response, and/or pharmacological outcome. The aim of these in vitro assays and in vivo markers is to support species selection but also to provide data to make qualitative and quantitative cross-species comparison to provide confidence that a model is capable of demonstrating potentially adverse consequences of target modulation and to support translational PK–PD strategies (ICH S6R).

By 2007, the tissue cross-reactivity assay (TCR) inadvertently was becoming, either from industry or regulatory creep (or both), the primary means to select species for safety assessment of monoclonal antibodies. The history, experience, methodology, and future directions of TCR studies in the development of antibody-based biopharmaceuticals are reviewed in Leach et al. (2010). The authors state that TCR studies are screening assays recommended for antibody and antibody-like molecules that contain a complementary determining region (CDR), primarily to identify off-target binding and secondarily to identify sties of on-target binding that were not previously identified. This was also the intent of both *step 4* of ICH S6 and the FDA Points to Consider document in the manufacture and testing of monoclonal antibody products for human use (FDA 1997). This intent is now reconfirmed in note 1 of the addendum: "TCR studies are in vitro tissue-binding assays employing immunohistochemical (IHC) techniques conducted to characterize binding of monoclonal antibodies and related antibody-like products to antigenic determinants in tissues. Other technologies can be employed in place of IHC techniques to demonstrate target/binding site distribution." The addendum also clarifies the value of TCR for species selection: "assessment of TCR in animal tissues is of limited value for species selection" (ICH S6R).

The technical difficulties regarding the conduct of TCR studies are recognized, and there is an acknowledgement that a clinical candidate may not be a good immunohistochemical (IHC) reagent and a TCR study might not always be technically feasible. Issues relating to the technical conduct and interpretation of TCR studies are reviewed in detail in Leach et al. (2010) and in publication based on an industry survey on the use of the TCR IHC assay (Bussiere et al. 2011).

The addendum purposely provides very little additional guidance on the use of alternative models such as transgenic models and homologous products over the ICH S6 guidance, except to state that such models can be considered when no relevant species can be identified. The use of animal models of disease to aid safety assessment is recommended when such models are used to evaluate proof of principle for monoclonal antibodies directed at foreign targets (i.e., bacterial, viral targets, etc.). Alternative approaches for toxicity testing of species-specific biopharmaceuticals still include animal models of disease, genetically modified mice, or use of homologous product (Bussiere et al. 2009; Bussiere 2008; Bornstein et al. 2009).

10.5.3 Duration of Studies

The addendum confirmed that the duration of repeat dose toxicity studies for chronic use products and 6-month duration in rodents and non-rodents are considered sufficient. The EWG reviewed published data and anonymized case studies provided by regulatory agencies and reached the view that toxicity studies of longer duration have not generally provided useful information that changed the clinical course of development in terms of altering clinical study design or patient information (Clarke et al. 2008).

10.5.4 "Tegenero"

Another key development in the field of preclinical safety assessment of biopharmaceuticals between 1997 and 2007 was the 2006 Tegenero incident with TGN1412, a superagonistic CD28-specific monoclonal antibody, in which six healthy human volunteers had to be admitted to a critical care unit during a first-in-human (FIH) study (Suntharalingham et al. 2006). Much has been published relating to this incident including commentary on best practice in nonclinical safety assessment, setting safe starting doses for first-in-human studies, the introduction of MABEL to reemphasize the importance of taking account of the pharmacologically active dose (PAD) as well as the NOAEL and HED, and the design of FIH studies (Schneider et al. 2006; Liedert et al. 2007; Horvath and Milton 2009; Milton and Horvath 2009; Lowe et al. 2009; Hansel et al. 2010). The incident also had an impact on industry/regulatory practice and regulatory guidance such as the publication in 2007 of the CHMP guideline on strategies to identify and mitigate risks for first-in-human clinical trials with investigations on medicinal products (EMEA/CHMP/SWP/28367/07). The implications of the incident were relevant for the ICH S6R(1) discussions in relation to the use of pharmacologically relevant species for safety assessment. The data made available to the public in the IMPD did not provide evidence that the cynomolgus monkey was a pharmacologically relevant species for the safety assessment of TGN1412: data on CD28 binding affinity for cynomolgus monkey was provided in the IMPD but apparently no data on in vitro functionality (e.g., T cell proliferation) nor was there evidence for in vivo pharmacological effects even at doses resulting in full target saturation. Furthermore, other relevant data with parental and surrogate TGN1412 molecules in humanized mouse models and rodents, and in vitro human data showing T cell proliferative activity, were not used in the overall safety assessment and safe starting dose selection (Horvath and Milton 2009).

Subsequent to the incident, new data demonstrated that white blood cells from cynomolgus monkeys do not respond to TGN1412 in the same way as human white blood cells, whether the cells are stimulated in vitro or in vivo. Essentially, TGN1412 is superagonistic in humans, but not in cynomolgus monkeys (Stebbings et al. 2007,

2009). Further work by the same group at NIBSC, UK, showed that activation of CD4+ effector memory T cells by TGN1412 was likely to be responsible for the cytokine storm experienced by the healthy volunteers. Furthermore, lack of CD28 expression on the CD4+ effector memory T cells of species used for preclinical safety testing of TGN1412 offers an explanation for the failure to predict a cytokine storm in humans (Eastwood et al. 2010). This illustrates the importance of understanding the target biology and mechanism of action of the biopharmaceutical product, the selection of pharmacologically relevant species for safety assessment, and also for an understanding of the limitations of the selected animal species for predicting safety for humans and where necessary supplementing these limitations by appropriate in vitro human systems to aid optimal selection of safe starting doses for FIH studies.

Experience with many monoclonal antibodies suggests that nonhuman primates appear not to predict cytokine release well for humans (Horvath and Milton 2009), and for this reason, the Tegenero incident triggered multiple workshops and publications relating to the development of in vitro human systems to predict cytokine release with the aim of addressing this limitation of nonhuman primates (Bugelski et al. 2009; Vidal et al. 2010; Findlay et al. 2011).

10.5.5 Dose Selection and Application of PK–PD Principles

An example of the industry-regulatory creep that was apparent by 2007, 10 years after S6 was finalized, was high dose selection for general toxicology studies. The intent of the S6 guidance was to allow sponsors to provide a scientific justification for dose selection, tailored to the specific product attributes, to achieve the aim of understanding pharmacological/physiological and toxicological dose response relationships in a pharmacologically relevant species. The guidance acknowledged the need for a case-by-case approach such that for some classes of products with little or no toxicity, it may not be possible to define a specific maximum dose, but for products with a lower affinity to or potency in the cells of the selected species than in humans cells, testing of higher doses may be important. By 2007, requests for sponsors to use the maximum tolerated dose (MTD) or maximum feasible dose (MFD) approaches were becoming more frequent suggesting a drift towards the small molecule approach where the use of such limit doses are common.

Over the last 10 years, many sponsors began applying pharmacokinetic–pharmacodynamic (PK–PD) modeling as an integral part of the preclinical and clinical development of protein drugs (Tabrizi and Roskos 2007; Tabrizi et al. 2009; Roskos et al. 2011). Greater emphasis was placed on translational strategies using bioanalytical data from appropriately selected and well-characterized PK and PD biomarker assays to allow a quantitative relationship between protein drug exposure, target modulation and biochemical, and physiological and pathophysiological effects to be established (Roskos et al. 2011). The selection of PD biomarkers that assess target engagement and modulation and downstream cellular effects can

provide proof of mechanism and also define the magnitude and duration of target modulation following drug administration. This PK–PD data can guide the selection of doses and dosing schedules for preclinical studies and clinical trials.

The S6R(1) addendum recognized the development of these translational PK–PD approaches and recommends the use of such approaches for high dose selection in general toxicity studies by identifying (1) a dose which provides the maximum intended pharmacological effect in the preclinical species and (2) a dose which provides an approximately tenfold exposure multiple over the maximum exposure to be achieved in the clinic. Following step 2 of S6R(1), the EWG received many comments requesting further clarification of the term "exposure," e.g., AUC, C_{max}, and $C_{average}$. However, the EWG decided to refrain from detailed guidance on this to allow sponsors to provide a scientific justification for the approach taken. The addendum also recognizes that appropriate PD endpoints are not always available, and in these cases, high dose selection can be based on PK data and available in vitro functional data.

10.5.6 Reproductive/Developmental Toxicity

The need for reproductive/developmental toxicity studies is dependent on the product, clinical indication, and intended patient population. The specific study design may be modified based on issues related to species specificity, immunogenicity, biological activity, and/or a long elimination half-life (ICH S6 Preclinical Safety Evaluation of Biotechnology-derived Pharmaceuticals 1997).

Both ICH S5A detection of toxicity to reproduction for medicinal products (ICHS5A Detection of Toxicity to Reproduction for Medicinal Products) and ICH S6 (ICH S6 Preclinical Safety Evaluation of Biotechnology-derived Pharmaceuticals 1997) allow flexible design strategies based upon scientific justification. The principles for assessing reproductive and developmental toxicity are guided by ICH S5A; the practices for biopharmaceuticals are guided by ICH S6. Selection of relevant species is critical to generating relevant risk information. Traditional species (rodents and rabbits) if relevant are preferred. A variety of animal models are acceptable for assessing reproductive/development effects of biopharmaceuticals homologous products that have also been used. Strategies vary based upon product attributes and intended use. Different strategies are also acceptable across similar product classes and indications (Cavagnaro 2010).

Nonhuman primates (NHP) are best used when the objective of the study is to characterize a relatively certain reproductive toxicant, rather than detect a hazard. According to ICH S5A, if it can be shown by means of kinetic, pharmacological, and toxicological data that the species selected is a relevant model for the human, a single species can be sufficient (ICHS5A Detection of Toxicity to Reproduction for Medicinal Products). Relevant measures of male fertility performance can be included in repeat dose toxicity studies if animals are sexually mature although assessing fertility is limited when using nonhuman primates.

The means by which biopharmaceuticals cross the placenta if at all may be species dependent considering the notable differences between rodent and primate placenta. For biopharmaceuticals that do not cross the placenta, embryo–fetal development (EF) studies in both rodents and NHP are likely to be restricted to maternal effects rather than direct teratogenic effects; thus, a study in rodents with a homologous product could probably model these effects as effectively as a study in primates (Martin and Weinbauer 2010).

10.5.6.1 Key Developments in Assessment of Reproductive/Developmental Toxicity Since ICH S6

The considerations in assessing the developmental and reproductive toxicity potential of biopharmaceuticals in traditional and nontraditional animal species are well summarized in an extensive review by Martin et al. (2009, 2010). This review provides a framework for developing DART testing strategies for biopharmaceuticals. In addition, it provides an overview of the state of DART testing by highlighting various strategies that have been implemented over the past two decades for approved biopharmaceuticals, the lessons learned, and the current challenges in the evaluation of novel biopharmaceuticals.

The guidance on DART testing was very abbreviated in ICH S6 and related mainly to study design issues and adaption of study designs which may be needed for biopharmaceuticals, rather than issues relating to species selection. When S6 was finalized in 1997, there were a few approved products for non-oncology indications which also showed species-restricted pharmacological activity such that the nonhuman primate was the only relevant species. The experience was limited to the interferons, some cytokines, and a few monoclonal antibodies. Since the finalization of S6, there has been an explosion in the development of products for which assessment of toxicity to reproduction is needed but for which the nonhuman primate is the only relevant species. As a result, the number of nonhuman primates used for reproductive toxicity testing was increasing dramatically (Martin et al. 2009; Chapman et al. 2009; Chellman et al. 2009).

This situation led to many questions and divergent regulatory scientific advice about the relative merits of the use of rodents versus non-rodent species such as nonhuman primates and the use of alternative models such as transgenics and homologous products in rodent reproductive toxicity studies. In addition, there were many questions about the optimal design of nonhuman primate studies to address questions relating to assessment of developmental and reproductive toxicity. These two areas were the main focus of the EWG discussions for the addendum.

Firstly, the EWG reconfirmed that the principles of developmental and reproductive toxicity (DART) testing for biopharmaceuticals are similar to those for small molecule pharmaceuticals and in general follow the regulatory guidance outlined in ICH S5(R2) (ICHS5A Detection of Toxicity to Reproduction for Medicinal Products). This includes the use of rodents and rabbits for embryo–fetal development studies with biopharmaceuticals if the clinical candidate is pharmacologically

active in both species, unless clear developmental toxicity has been identified in one species. Several regulatory regions stated during the preparation of the S6 addendum that this requirement for two species for embryo–fetal development (EFD) studies was based on a review of internal databases and product labels and a lack of justification for the use of only a rodent or a rabbit.

An aspect which was considered by the EWG was the placental transfer of biopharmaceuticals. Small molecules (<1,000 Da) and their metabolites can diffuse across plasma membranes and the placenta by simple diffusion. In contrast, large molecule biopharmaceuticals do not appreciably diffuse across plasma membranes, including the placenta, and, therefore, have limited access to the conceptus. However, certain types of large molecules, such as monoclonal antibodies, can cross the placenta in mid- and late gestation by Fc receptor-mediated endocytosis via FcRn receptors (Martin et al. 2009; Simister 2003).

There are also species differences in placental transfer of antibodies between rodents and primates (Martin et al. 2009; Pentsuk and van der Laan 2009). In humans and nonhuman primates, transfer of antibodies across the placenta occurs primarily during the latter part of pregnancy, i.e., after organogenesis. This also appears to be the case for rabbits. In contrast, in rodents, transfer across the visceral yolk sac begins earlier in pregnancy, permitting exposure during organogenesis. Consequently, rodents may overestimate the human risk. However, the available data for some species is rather old and relates mainly to endogenous immunoglobulins induced by immunization to various antigens. The BioSafe group is in the process of gathering available data on placental transfer on a wide range of antibody and antibody-related products in development and plan to identify and fill data gaps to enable a better understanding of species differences in placental transfer.

One conclusion from the available information on the pattern of placental transfer in humans is that study designs that allow the detection of both indirect effects in early gestation plus the effects of direct fetal exposure in mid- and late gestation are recommended for developmental toxicity of monoclonal antibodies and related products.

There are increasing numbers of reports, many so far unpublished, of treatment-related fetal anomalies with monoclonal antibodies when administered to nonhuman primates only during the period of major organogenesis. One such published report related to figitumumab, an anti-insulin-like growth factor-1 receptor (IGF-1R) (Bowman et al. 2010). Thus, even low-level placental transfer and embryo–fetal exposure to potent monoclonal antibodies in early gestation may be sufficient to result in developmental toxicity.

Over several EWG meetings on the addendum, the regulators expressed a preference for DART testing with the clinical candidate, even if the only relevant species is a nonhuman primate. The EWG recognized the difficult balance between the limitations of a study in nonhuman primates with the clinical candidate versus the greater power of rodent developmental and reproductive toxicity studies but using a homologous product. However, although a preference is stated in the addendum, this does not mean that use of the nonhuman primate is the only acceptable option and a sponsor may still be able to provide a scientific justification for the use of an alternative DART testing strategy such as the use of alternative models including the use of homologous

products in rodent studies. This justification is likely to be based on the value of such alternative approaches to the communication and management of risk to humans.

There is now widespread industry and regulatory acceptance of the enhanced pre- and postnatal study (ePPND) design option when using nonhuman primates. This ePPND study combines the traditional "segmented" EFD study with the pre- and postnatal development (PPND) study into a single "enhanced" PPND study design where a single cohort of nonhuman primates is exposed throughout gestation and allowed to give birth naturally (Stewart 2009). The proposed "enhanced" PPND study design evaluates all the stages of the traditional two-study design using fewer animals. It also assesses the functional consequences of mid- to late gestational exposure (Martin and Weinbauer 2010; Chellman et al. 2009). This is of particular relevance to the risk assessment of monoclonal antibodies where fetal exposure to maternal IgG increases as pregnancy progresses and where morphologic examination of a preterm fetus may not be adequate to reveal the presence of adverse effects on functional development of key target organs. Another topic of hot debate in the EWG was the numbers of animals to be used in nonhuman primate ePPND studies. ICH S5(R2) note 13 states that for all but the rarest events (such as malformations, abortions, total litter loss), evaluation of between 16 and 20 litters for rodents and rabbits tends to provide a degree of consistency between studies. However, the same note also acknowledges that there is very little scientific basis underlying specified group sizes in past and existing guidelines nor in S5(R2). The numbers specified are educated guesses governed by the maximum study size that can be managed without undue loss of overall study control. The use of nonhuman primates carries additional ethical concerns, but number of animals per group should still be sufficient to allow meaningful interpretation of the data. An evaluation of pregnancy and infant loss in 1,069 vehicle-treated cynomolgus monkeys from 78 EFD studies and 14 PPND studies accrued during 1981–2007 was reported by Jarvis et al. (2010) to review the variability of pregnancy losses and impact on statistical power estimates and group size considerations. This evaluation indicated that based on the variability of pregnancy losses in this database and that in a PPND study with initial vehicle-control group sizes of 16 or 20, there is an 80 % likelihood of having 9 or 11 infants at day 7 postpartum, respectively.

After long debates on this topic, the addendum now states that "developmental toxicity studies in NHPs can only provide hazard identification. The number of animals per group should be sufficient to allow meaningful interpretation of the data (see Note 5)" (ICH S6R). Note 5 b refers to Jarvis et al. (2010) and recommends that group sizes in ePPND studies should yield a sufficient number of infants (6–8 per group at postnatal day 7) in order to assess postnatal development (Jarvis et al. 2010).

The addendum also outlines possibilities to reduce nonhuman primate use still further, e.g., by the use of fewer treatment groups (Chapman et al. 2012), reuse of vehicle-control maternal animals, early termination of animal accrual into the study if a treatment-related effect is noted during the course of the study, and use of a limited number of animals to confirm a likely hazard based on cause for concern based on mechanism of action (note: a study in rodents with a homologous product may also be justifiable in this case).

The evaluation of fertility is also problematic in nonhuman primates, and the addendum recognizes that mating studies are not practical for NHPs. Nonhuman primates are similar to humans with respect to the physiology and endocrinology of testicular and ovarian function (Chellman et al. 2009; Weinbauer et al. 2008), and potential effects on male and female fertility can be assessed by evaluation of the reproductive tract (organ weights and histopathological evaluation) in studies of at least 3-month duration using sexually mature nonhuman primates. The intent of the addendum was that the evaluation of potential effects on fertility in sexually mature nonhuman primates would be combined with the evaluation of general toxicity, usually the evaluation of chronic toxicity. Additional endpoints such as menstrual cyclicity, sperm counts, sperm morphology/motility, and male and female reproductive hormone levels are recommended if there is a specific cause for concern based on pharmacological activity or previous findings. Menstrual cyclicity is a fairly easy endpoint to monitor in cynomolgus monkeys by daily vaginal smears, and many sponsors chose to include this endpoint routinely in such studies rather than "for cause." However, the practical and logistical issues need to be recognized in order to get meaningful menstrual cyclicity data. When using social-housed female cynomolgus monkeys, it is essential to consider the housing history and familiarity between the animals prior to pair or group formations since this can lead to irregular cyclicity (Weinbauer et al. 2008). There is a need for several months of pre-study acclimation to the facility and cage mates.

Because mating studies are not practical for NHPs, there is a "data gap" in relation to a lack of information on the effects on conception and implantation. The addendum recommends that this data gap is addressed in several ways (1) experimentally using a homologous product in rodent studies or (2) risk mitigation through clinical trial management procedures, informed consent, and appropriate product labeling. It is not recommended to produce a homologous product or alternative animal model solely to conduct mating studies in rodents and to fill this data gap on effects on conception and implantation.

The timing of assessment of developmental and reproductive toxicity during clinical development was also a main topic for discussion in the EWG, in parallel to discussions ongoing in the ICH M3(R2) EWG. Both S6R(1) and M3(R2) recognize the difficulty of conducting developmental toxicity studies in nonhuman primates when this species is the only relevant species and allow for the conduct of such studies during phase III, providing there are sufficient precautions to prevent pregnancy, and the lack of animal reproductive toxicity data is communicated in the informed consent (ICH M3(R2), 2010).

Overall, while the addendum does express a preference for developmental and reproductive toxicity testing of the clinical candidate, various possibilities are suggested for reducing the overall number of monkeys in the reproductive toxicity testing strategy if use of the nonhuman primate is the only option for such testing. The use of homologous products in rodent studies rather than testing of the clinical candidate may also be appropriate where there is adequate scientific justification provided by the developer for the DART strategy proposed.

10.5.7 Genotoxicity

Genotoxicity testing is routinely conducted for pharmaceuticals to detect mutagenic and clastogenic compounds that may be carcinogens. Assays are designed to detect mutagenicity and clastogenicity, but not cellular proliferation. While uptake of low-molecular-weight compounds occurs through passive diffusion or nonspecific pinocytosis, large-molecular-weight compounds require active transport. Specific transporter mechanisms are typically not present in current assay systems thus "not relevant models" for assessing biopharmaceuticals (Cavagnaro 2010). False positives have been observed in the standard Ames test due to the presence of growth-promoting constituents in the test samples such as histidine or its precursors. Positive results have also been shown for lipase, glucagon, erythropoietin, and DNAse presumably based upon pharmacological activity hence considered predictable as exaggerated pharmacology.

While studies may be applicable for protein conjugates with a chemical organic linker, consideration is warranted particularly when a residual organic linker is found in the product because of the instability of the conjugate during storage or upon dilution in the serum. Additionally, unlike pharmaceuticals where there may be a cause for concern for testing impurities for potential genotoxic potential, impurities associated with biopharmaceuticals are generally referred to as process related and include residual host cell proteins, fermentation components, column leachables, and detergents rather than organic chemicals and as such not considered to pose mutagenic risks.

Biopharmaceuticals do not have the same distribution properties as small molecules and are therefore not expected to pass through cell and nuclear membranes to interact with DNA. Experience has confirmed that the standard battery of genotoxicity assays is not relevant for products that do not directly interfere with DNA or mitosis to induce gene mutations, chromosome aberrations, or DNA damage. While studies may be applicable for protein conjugates with a chemical organic linker, consideration is warranted if there is precedence of use with the linker or if there is no evidence of degradation of the protein conjugate. Additionally, unlike small molecules where there may be a cause for concern for testing for genotoxic impurities, process-related impurities associated with biopharmaceuticals include residual host cell proteins, fermentation components, column leachables, and detergents rather than organic chemicals.

10.5.7.1 Key Developments in Assessment of Genotoxicity Since ICH S6

Experience confirmed that the standard battery of genotoxicity assays is not relevant for products that do not directly interfere with DNA or mitosis to induce gene mutations, chromosome aberrations, or DNA damage. In a retrospective review of 78 compounds, mostly recombinant peptides and proteins, Gocke et al. (1999) concluded that genotoxicity testing of biological drugs was generally inappropriate and unnecessary.

10.5.8 Carcinogenicity

Carcinogenicity studies in two species are generally required for pharmaceuticals administered chronically. The need for carcinogenicity assessment of a biopharmaceutical is determined by a number of factors and is similar to those for pharmaceuticals. However, most of the early biotechnology molecules developed were for severe clinical indications and/or addressed unmet medical needs.

In cases where a biopharmaceutical is active and relatively non-immunogenic in rodents, and studies have not provided sufficient information to allow an assessment of carcinogenic potential, then a single bioassay has been considered per ICH S6 (e.g., a 2-year bioassay was performed for DNAse due to the mechanism of action and intended patient population). However, the standard bioassay was generally considered irrelevant for biopharmaceuticals (ICH S6 Preclinical Safety Evaluation of Biotechnology-derived Pharmaceuticals 1997). One reason is that molecular structure excludes biopharmaceuticals from being intrinsically carcinogenic and as mentioned above, there would not be a concern for potential "carcinogenic metabolites." In addition, the rodent bioassay may otherwise not be relevant based on a high degree of antibody formation following repeat dosing of the clinical candidate, the lack of availability of an alternate product (e.g., homologous protein, surrogate molecule), or the lack of sufficient comparability.

ICH S6 guidance recommended incorporation of sensitive indices of cellular proliferation in chronic dose toxicity studies. However, it is recognized that while qualitative or quantitative increases in proliferation of target tissue and increases in organ weight signaling preneoplastic changes may represent early signals of epigenetic mechanisms, not all hyperplasia will result in neoplasia.

10.5.8.1 Key Developments in Assessment of Carcinogenicity Since ICH S6

The past and current practice over the last two decades regarding carcinogenicity assessments of biopharmaceuticals was reviewed by a collaborative effort of industry toxicologists involved in the preclinical development of biopharmaceuticals (Vahle et al. 2010). This review includes publicly available information on 80 approved biopharmaceuticals. No assessments related to carcinogenicity or tumor growth promotion were identified for 51 of the 80 molecules. For the 29 biopharmaceuticals in which assessments related to carcinogenicity were identified, various experimental approaches were employed. The review concluded that the traditional 2-year carcinogenicity assays should not be considered the default method for biopharmaceuticals and that if experimentation is considered warranted, it should be hypothesis driven and may include a variety of experimental models. Ultimately, it is important that preclinical data provide useful guidance in product labeling.

In parallel to the EWG discussion on assessment of carcinogenic potential, the value of the 2-year rodent bioassay for predicting carcinogenic hazard for humans of pharmaceutical products was also under review (Sistare et al. 2010; Friedrich and

Olejniczak 2010). Carcinogenicity data for pharmaceuticals and biopharmaceuticals approved via the European centralized procedure between 1995 and 2009 were evaluated; 65 % of compounds were deemed positive for carcinogenicity in at least one long-term carcinogenicity study or in repeat dose toxicity studies (Friedrich and Olejniczak 2010). These authors concluded that "due to the high number of rodent tumor findings with unlikely relevance for humans, the value of the currently used testing strategy for carcinogenicity appears questionable. A revision of the carcinogenicity testing paradigm is warranted." A pharmaceutical industry group made a proposal to refine regulatory criteria for conducting a 2-year rat study with pharmaceuticals to be based on assessment of histopathological findings from a rat 6-month study, evidence of hormonal perturbation, genetic toxicology results, and the findings of a 6-month transgenic mouse carcinogenicity study (Sistare et al. 2010).

Bugelski et al. (2010) reviewed the preclinical approaches to evaluate the potential of immunosuppressive drugs to influence human neoplasia. The authors concluded that the 2-year rodent bioassay performs poorly in identifying the mechanism of action-related hazard for developing certain tumor types, especially lymphomas and skin cancer. Classifying immunosuppressive drugs based on their mechanism of action and hazard identification from preclinical studies and a prospective pharmacovigilance program to monitor carcinogenic risk was proposed as a feasible way to manage patient safety during the clinical development program and post-marketing.

At the first EWG meeting for the addendum in 2008, there was a recognition that the issues encountered regarding the assessment of carcinogenic potential of biopharmaceuticals were likely related to the industry-regulatory creep and changing regulatory environment alluded to earlier. These issues were likely related more to implementation of the S6 guidance in some regulatory regions rather than lack of clarity of this guidance.

The S6 guidance started from the general philosophy that standard carcinogenicity bioassays are generally inappropriate for biopharmaceuticals but that a product-specific assessment of carcinogenicity may still be needed. By 2007, the general philosophy of some regulatory agencies was the same as for small molecules—"if you can do it, you should do it"—if such an assessment is needed according to the clinical population and treatment duration (ICH S1A).

The EWG reviewed the practice of carcinogenicity testing of biopharmaceuticals over the last two decades and also reviewed several case studies provided by some regulatory agencies. Overall, the general philosophy as outlined in the S6 guidance was upheld, and attempts were made to clarify certain aspects. When an assessment of carcinogenic potential is warranted, it is up to the sponsor to design a strategy to address the potential hazard, based on a weight of evidence approach and an understanding of target biology related to potential carcinogenic concern. Rodent bioassays (or short-term carcinogenicity studies) with homologous products were generally considered to be of limited value to assess carcinogenic potential of the clinical candidate. Ultimately, the product-specific assessment of carcinogenic potential is used to communicate risk and provide input to the risk management plan along with labeling proposals, clinical monitoring, post-marketing surveillance, or a combination of these approaches (Cavagnaro 2008b).

10.5.9 Immunogenicity

ICH S6 states, "Most biotechnology-derived pharmaceuticals intended for humans are immunogenic in animal" (ICH S6 Preclinical Safety Evaluation of Biotechnology-derived Pharmaceuticals 1997). Traditional antigenicity studies or guinea pig anaphylaxis studies are not useful for predicting immunogenicity in humans and are now generally recognized as not being appropriate studies for biologics. When these studies were conducted with biopharmaceuticals, they were not surprisingly positive and led to adverse effects in animals. Since there is little to no predictive value in these studies, and they were not considered appropriate, such studies have not been conducted since publication of ICH S6.

Administration of human proteins in sufficient quantity into animals is expected to elicit an immunological response. Even homologous/surrogate molecules have induced immune responses in the respective species. Immunogenicity assessments are conducted to assist in the interpretation of the study results and design of subsequent studies rather than to predict potential immunogenicity in humans. The presence of neutralizing antibody can change the PK/PD profile and thus impact exposure margins and estimates of toxicity. In early studies with biopharmaceuticals, the development of antibodies in a toxicology study was considered a reason to stop studies; however, we now know that we can "dose through" in animals similar to dosing practices in humans. While the presence of antibodies in animals is generally not predictive for humans, the information has helped in defining relative immunogenicity and in identifying potential consequences of an immune response, e.g., neoantigenicity, autoantigenicity, immune complex deposition, complement activation, and the impact of antibodies crossing the placenta.

The two major areas of concern relating to the assessment of antigenic/immunogenic potential are (1) product/active ingredient and (2) process/excipient/final formulation. The formation of antibodies is monitored at various intervals throughout toxicity studies in order to be able to interpret the studies and determine if there is any impact on exposure. Information should be provided on the effect of antibody formation on the pharmacokinetic behavior of the product and whether antibodies interfere with the assay used to monitor the product in biological fluids. Clinically relevant antibodies include clearing antibodies, sustaining antibodies, neutralizing antibodies, and antibodies that cross-react with endogenous proteins. The presence of neutralizing antibodies and abrogation of subsequent pharmacological and/or toxicological effects can provide the justification for limiting the duration of repeated dose studies. However, the presence of antidrug antibodies in the absence of PK effects, neutralization of activity, or other toxicities is not sufficient to support study termination or shorter study durations.

10.5.9.1 Key Developments in Assessment of Immunogenicity Since ICH S6

By 2007, it had become apparent that immunogenicity testing was being largely driven by bioanalytical considerations with great emphasis being given to the S6 guidance

that "measurement of antibodies…should be performed when conducting repeated dose toxicity studies …" and "antibody responses should be characterized (e.g. titre, number of responding animals, neutralizing or non-neutralizing)" (ICH S6 Preclinical Safety Evaluation of Biotechnology-derived Pharmaceuticals 1997). The primary purpose of such immunogenicity testing in support of toxicity studies "in order to aid in the interpretation of these studies" seemed to be superseded by bioanalytical considerations. Because of assay sensitivity issues relating to drug interference, the perceived requirement to measure and characterize antibody responses in repeated dose toxicity studies in order to determine whether an animal was antidrug antibody (ADA) positive or negative was driving long treatment-free recovery periods, even in the absence of any toxicity findings needing an evaluation of reversibility.

A decision tree for conducting ADA analyses to support nonclinical study interpretation was provided by Ponce et al. (2009). This decision tree is intended to guide the investigator through a series of considerations to determine whether ADA analysis is necessary to aid in the interpretation of a study. The authors concluded that immunogenicity data should be integrated with available clinical and anatomic pathology, PK, and PD data to properly interpret nonclinical studies. PD markers of target engagement such as ligand capture (soluble ligand) or receptor occupancy (cell surface ligand), as well as downstream signaling markers or other in vivo mechanistic markers, also contain valuable information regarding the neutralizing potential of an ADA response evident as loss of target engagement or loss of functional or pharmacological activity. Where such PD markers are available, the need for specific neutralization assays may be obviated by the use of these alternative markers of functional activity (Buttel et al. 2011).

The S6R(1) addendum clarifies the purpose of immunogenicity testing in the first sentence: "immunogenicity assessments are conducted to assist in the interpretation of the study results and design of subsequent studies." The addendum provides clarification for when measurement of antidrug antibodies (ADA) in nonclinical studies should be evaluated and when characterization of neutralization potential is warranted. When no PD marker exists to demonstrate sustained activity in the in vivo toxicology studies, characterization of neutralizing potential is warranted, but the addendum provides clarification that this can be assessed indirectly with an ex vivo bioactivity assay or an appropriate combination of assay formats for PK–PD (Buttel et al. 2011) or directly in a specific neutralizing antibody assay.

10.6 Conclusions

Preclinical safety evaluation of biopharmaceuticals has evolved through the application of scientific insight, historical and anecdotal experiences, and common sense. The scientific community has relied on the exchange of ideas between academia, industry, and regulatory scientists. Many new challenges in biopharmaceutical clinical development lie ahead. New technologies and products not yet envisioned will continue to

challenge toxicologists. Additional challenges and advances will come from efforts devoted to site-directed delivery or site-specific expression. Open dialogue between scientists who are regulators, academic scientists, or industry scientists will be critical in ensuring that the new products that are safe and effective are made available without unnecessary delay. A regulatory environment that encourages innovation will make this possible.

Development practices for preclinical safety assessment of biopharmaceuticals have been and will continue to be a dynamic process that is strongly controlled by the expanding knowledge and the innovations in product design. However, the full investigation of the potential usefulness of biopharmaceuticals will require the development of reliable animal model systems that allow assessment of toxicity and provide pharmacokinetic data that can be successfully scaled to humans in order to reduce risk factors before clinical testing. There is also a need to develop and refine appropriate human in vitro systems to aid safety assessment in cases where reliable animal models do not exist but also to address specific limitations of animal studies, e.g., assessing the potential for cytokine release (Vidal et al. 2010). Once sufficient data have accrued, it is important to review experiences as was done in the case of the ICH S6 and recalibrate approaches if necessary.

The design of relevant preclinical safety evaluation programs is consistent with global initiatives to facilitate and to improve clinical development programs. In the coming years, stakeholders will be facing the issue of how to implement preclinical development programs for biopharmaceuticals and pharmaceuticals that better anticipate adverse effects including development of new test systems that produce reliable results faster and at lower cost. Hopefully, preclinical evaluation programs will evolve and mature concurrently with more novel products and will focus on improving the predictive value of preclinical safety testing, challenging toxicologists to provide information from the most appropriate studies.

Biotechnology has provided not only the hope of potential new therapies but also the necessary tools to evaluate new therapies. Toxicology as a science has benefited from this experience in many ways. The case-by-case approach to preclinical safety evaluation should continue to provide for scientific advancement in toxicology and the inducement of quality research into relevant safety assessment for the next generation of novel therapies.

References

Bass R, Kleeburg U, Schroder H et al (1992) Current guidelines for the preclinical safety assessment of therapeutic proteins. Toxicol Lett 64/65:339–347

Bornstein GG, Klakamp SL, Andrews L, Boyle WJ, Tabrizi M (2009) Surrogate approaches in development of monoclonal antibodies. Drug Discov Today 14(23/24):1159–1165

Bowman CJ, Chmielewski G, Oneda S, Finco D, Boucher MA, Todd M (2010) Embryo-fetal development toxicity of figitumumab, an anti-insulin-like growth factor-1 receptor (IGF-1R) monoclonal antibody, in cynomolgus monkeys. Birth Defects Res B Dev Reprod Toxicol 89(4):326–338

Brennan FR, Shaw L, Wing MG, Robinson C (2004) Preclinical safety testing of biopharmaceuticals. Mol Biotechnol 27:59–74

Buckley LA, Benson K, Davis-Bruno K, Dempster M, Finch GL, Harlow P, Haggerty HG, Hart T, Kinter L, Leighton JK, McNulty J, Roskos L, Saber H, Stauber A, Tabrizi M (2008) Nonclinical aspects of biopharmaceutical development: discussion of case studies at a PhRMA-FDA workshop. Int J Toxicol 27(4):303–312

Bugelski PJ, Achuthanandam R, Capocasale RJ, Treacy G, Bouman-Thio E (2009) Monoclonal antibody-induced cytokine release syndrome. Expert Rev Clin Immunol 5(5):499–521

Bugelski P, Volk A, Walker MR, Krayer JH, Martin P, Descotes J (2010) Critical review of preclinical approaches to evaluate the potential of immunosuppressive drugs to influence human neoplasia. Int J Toxicol 29(5):435–466

Bussiere J (2008) Species selection considerations for preclinical toxicology studies for biotherapeutics. Expert Opin Drug Metab Toxicol 4(7):871–877

Bussiere JL, Martin P, Horner M, Couch J, Flaherty M, Andrews L, Beyer J, Horvath C (2009) Alternative strategies for toxicity testing of species-specific biopharmaceuticals. Int J Toxicol 28:230–253

Bussiere JL, Leach MW, Price KD, Mounho BJ, Lighfoot-Dunn R (2011) Survey results on the use of the tissue cross-reactivity immunohistochemistry assay. Regul Toxicol Pharmacol 59:493–502

Buttel IC, Chamberlain P, Chowers Y et al (2011) Taking Immunogenicity assessment of therapeutic proteins to the next level. Biologicals 39:100–109

Cavagnaro JA (1992a) Misconceptions within biotechnology. In: D'Arcy PF, Harmon DWG (eds) Proceedings of the first conference on harmonisation, Brussels, 1991. The Queen's University of Belfast, Belfast, pp 301–307

Cavagnaro JA (1992b) Science-based approach to preclinical safety evaluation of biotechnology products. Pharmaceut Eng 12:32–33

Cavagnaro JA (2002) Preclinical safety evaluation of biotechnology-derived pharmaceuticals. Nat Rev Drug Discov 1:469–475

Cavagnaro JA (ed) (2008a) Preclinical safety evaluation of biopharmaceuticals: a science-based approach to facilitating clinical trials. Wiley, New York

Cavagnaro JA (2008b) Assessment of carcinogenic risk of biopharmaceuticals. In: Cavagnaro JA (ed) Preclinical safety evaluation of biopharmaceuticals: a science-based approach to facilitating clinical trials. Wiley, New York, pp 399–477

Cavagnaro J (2010) Considerations for the preclinical safety evaluation of biopharmaceuticals. In: Charlene A (ed) McQueen, comprehensive toxicology, vol 3. Academic, Oxford, pp 29–51

Chapman K, Pullen N, Coney L, Dempster M, Andrews L, Bajramovic J, Baldrick P, Buckley L, Jacobs A, Hale G, Green C, Ragan I, Robinson V (2009) Preclinical development of monoclonal antibodies: considerations for the use of non-human primates. mAbs 1(5):505–516

Chapman KL, Andrews L, Bajramovic JJ, Baldrick P, Black LE, Bowman CJ, Buckley LA, Coney LA, Couch J, Dempster AM, De Haan L, Jones K, Pullen N, de Boer AS, Sims J and Ragan CI (2012) The design of chronic toxicology studies of monoclonal antibodies: Implications for the reduction in the use of non-human primates. Regul Toxicol Pharmacol 62(2):347–354

Chellman GJ, Bussiere JL, Makori N, Martin PL, Ooshima Y, Weinbauer GF (2009) Developmental and reproductive toxicology studies in nonhuman primates. Birth Defects Res B 86:446–462

Clarke J, Hurst C, Martin P, Vahle J, Ponce R, Mounho B, Heidel S, Andrews L, Reynolds T, Cavagnaro J (2008) Duration of chronic toxicity studies for biotechnology-derived pharmaceuticals: is 6 months still appropriate? Regul Toxicol Pharmacol 50:2–22

Claude JR (1992) Difficulties in conceiving and applying guidelines for the safety evaluation of biotechnologically-produced drugs: some examples. Toxicol Lett 64(65):349–355

Dayan AD (1987) Rationality and regulatory requirements- a view from Britain. In: Grahm CE (ed) Preclinical safety of biotechnology products intended for human use: clinical and biological research, vol 235. Alan R. Liss, New York, pp 89–106

Dayan AD (1995) Safety evaluation of biological and biotechnology-derived medicines. Toxicology 105:59–68

Eastwood D, Findlay L, Poole S, Bird C, Wadhwa M, Moore M, Burns C, Thorpe R, Stebbings R (2010) Monoclonal antibody TGN1412 trial failure explained by species differences in CD28 expression on $CD4^+$ effector memory T-cells. Br J Pharmacol 161:512–526

EMEA/CHMP/SWP/28367/07

FDA (1997) Points to consider in the manufacture and testing of monoclonal antibody products for human use, 1997. http://www.fda.gov

Findlay L, Eastwood D, Ball C, Robinson CJ, Bird C, Wadhwa M, Thorpe R, Stebbings R, Poole S (2011) Comparison of novel methods for predicting the risks of pro-inflammatory clinical infusion reactions during monoclonal antibody therapy. J Immunol Methods 371(1–2):131–142

Finkle BS (1987) Genetically engineered drugs: toxicology with a difference. In: Grahm CE (ed) Preclinical safety of biotechnology products intended for human use: clinical and biological research, vol 235. Alan R. Liss, New York, pp 161–167

Friedrich A, Olejniczak K (2010) Evaluation of carcinogenicity studies for medicinal products for human use authorized via the European centralized procedure (1995–2009). Regul Toxicol Pharmacol 60(2):225–248

Galbraith WM (1987) Symposium discussion. In: CE Grahm (ed) Preclinical safety of biotechnology products intended for human use: clinical and biological research, vol 235. Alan R. Liss, New York, pp 189–206

Giss HE (1987) Foreword. In: Grahm CE (ed) Preclinical safety of biotechnology products intended for human use: clinical and biological research, vol 235. Alan R. Liss, New York, pp xiii–xv

Gocke E, Albertini S, Brendel-Schwaab S, Muller L, Suter W et al (1999) Genotoxicity testing of biotechnology-derived products. Report of GUM task force. Mutant Res 436:137–156

Griffith SA, Lumley CE (1998) Non-clinical safety studies for biotechnologically-derived pharmaceuticals: conclusions form an international workshop. Hum Exp Toxicol 17:63083

Hansel TT, Kropshofer H, Singer T, Mitchell JA, George AJT (2010) The safety and side effects of monoclonal antibodies. Nat Rev Drug Discov 9:325–338

Hayakawa T (1992) Current regulatory situation in Japan with respect to the preclinical safety testing for biotechnology products intended for human use. In: D'Arcy PF, Harmon DWG (eds) Proceedings of the first conference on harmonisation, Brussels 1991. The Queen's University of Belfast, Belfast, pp 296–301

Hayes TJ, Cavagnaro JA (1992) Progress and challenges in the preclinical assessment of cytokines. Toxicol Lett 64(65):291–297

Henck JW, Hilbish KG, Serabian MA, Cavagnaro JA, Hendrickx AG, Agnish ND, Kung ADH, Mordenti J (1996) Reproductive toxicity testing of therapeutic biotechnology agents. Teratology 53:185–195

Hohbach C (1992) Safety in biotechnology progress and problems, industrial point of view. In: D'Arcy PF, Harmon DWG (eds) Proceedings of the first conference on harmonisation, Brussels 1991. The Queen's University of Belfast, Belfast, pp 307–316

Horvath CJ, Milton MN (2009) The TeGenero incident and the Duff Report conclusions: a series of unfortunate events or an avoidable event? Toxicol Pathol 37:372–383

ICH M3(R2) (2010): Guidance on nonclinical safety studies for the conduct of human clinical trials and marketing authorization for pharmaceuticals. http://www.ich.org

ICH S6 Preclinical safety evaluation of biotechnology-derived pharmaceuticals (1997) http://www.ich.org

ICH S6R(1): Preclinical safety evaluation of biotechnology-derived pharmaceuticals. http://www.ich.org

ICHS5A Detection of toxicity to reproduction for medicinal products

Jarvis P, Srivastav S, Vogelwedde E, Stewart J, Mitchard T, Weinbauer GF (2010) The cynomolgus monkey as a model for developmental toxicity studies: variability of pregnancy losses, statistical power estimates, and group size considerations. Birth Defects Res B 89:175–187

Kikuchi Y (1992) Panel presentation safety: biotechnology. In: D'Arcy PF, Harmon DWG (eds) Proceedings of the first conference on harmonisation, Brussels 1991. The Queen's University of Belfast, Belfast, pp 316–319

Leach MW, Halpern WG, Johnson CW, Rojko JL, MacLachlan TK, Chan CM, Galbreath EJ, Ndifor AM, Blanset DL, Polack E, Cavagnaro JA (2010) Use of tissue cross-reactivity studies in the development of antibody-based biopharmaceuticals; history, experience, methodology, and future directions. Toxicol Pathol 38:1138–1166

Liedert B, Bassus S, Schneider CK, Kalinke U, Lower J (2007) Safety of Phase I clinical trials with monoclonal antibodies in Germany – the regulatory requirements viewed in the aftermath of the TGN1412 disaster. Int J Clin Pharmacol Ther 45(1):1–9

Lowe P, Tannenbaum S, Wu K, Lloyd P, Sims J (2009) On setting the first dose in man: quantitating biotherapeutic drug-target binding through pharmacokinetic and pharmacodynamics models. Basic Clin Pharmacol Toxicol 106:195–209

Martin PL, Breslin W, Rocca M, Wright D, Cavagnaro J (2009) Consideration in assessing the developmental and reproductive toxicity potential of biopharmaceuticals. Birth Defects Res B 86:176–203

Martin PL, Weinbauer GF (2010) Developmental toxicity testing of biopharmaceuticals in nonhuman primates: previous experience and future directions. Int J Toxicol 29:552–568

Milton MN, Horvath CJ (2009) The EMEA guideline on first-in-human clinical trials and its impact on pharmaceutical development. Toxicol Pathol 37:363–371

Nakazawa T, Kai S, Kawai M, Maki E, Sagami F, Onodera H, Kitajima S, Inoue T (2004) "Points to Consider" regarding safety assessment of biotechnology-derived pharmaceuticals in nonclinical studies (English Translation). J Toxicol Sci 29:497–504

Pentsuk N, van der Laan JW (2009) An interspecies comparison of placental antibody transfer: new insights into developmental toxicity testing of monoclonal antibodies. Birth Defects Res B 86:328–344

Ponce R, Abad L, Amaravadi L, Gelzieichter T, Gore E, Green J, Gupta S, Herzyk D, Hurst C, Ivens I, Kawabata T, Mauer C, Mounho B, Rup B, Shankdar G, Smith H, Thomas P, Wierda D (2009) Immunogenicity of biotechnology-derived therapeutics: assessment and interpretation of nonclinical studies. Regul Toxicol Pharmacol 54:164–182

Roskos L, Schneider A, Vainshtein I, Schwickart M, Lee R, Lu H, Faggioni R, Liang M (2011) PK-PD modeling of protein drugs: implications in assay development. Bioanalysis 3(6):659–675

Ryan AM, Terrell TG (2002) Biotechnology and its products. In: Handbook of toxicologic pathology, 2nd edn. Academic, New York, pp 479–500

Schneider CK, Kalinke I, Lower J (2006) TGN1412 – a regulator's perspective. Nat Biotechnol 24(4):368

Serabian MA, Pilaro AM (1999) Safety assessment of biopharmaceuticals: ICH and beyond. Toxicol Pathol 27:27–31

Simister N (2003) Placental transport of immunoglobulin G. Vaccine 21:3365–3369

Sims J (2001) Assessment of biotechnology products for therapeutic use. Toxicol Lett 120:59–66

Sistare FD, Morton D, Alden C, Christensen J, Keller D et al (2010) An analysis of pharmaceutical experience with decades of rat carcinogenicity testing: support for a proposal to modify current regulatory guidelines. Toxicol Pathol 39(716–744):2011

Stebbings R, Findly L, Edwards C, Eastwood D, Bird C, North D, Mistry Y, Dilger P, Liefooghe E, Cludts I, Fox B, Tarrant G, Robinson J, Meager T, Dolman C, Thorpe SJ, Bristow A, Wadhwa M, Thorpe R, Poole S (2007) "Cytokine storm" in a Phase I trial of monoclonal antibody TGN1412: better understanding the causes to improve the preclinical testing of immunotherapeutics. J Immunol 179:3325–3331

Stebbings R, Pool S, Thorpe R (2009) Safety of biologics, lessons learnt from TGN1412. Curr Opin Biotechnol 20(6):673–677

Stewart J (2009) Developmental toxicity testing of monoclonal antibodies: an enhanced pre-and postnatal study design option. Reprod Toxicity 28:220–225

Suntharalingham G, Perry MR, Ward S, Brett SJ, Castello-Cortes A, Brunner MD, Panoskaltsis N (2006) Cytokine storm in a Phase I trial of the anti-CD28 monoclonal antibody TGN1412. N Engl J Med 355:1018–1028

Tabrizi MA, Roskos LK (2007) Preclinical and clinical safety of monoclonal antibodies. Drug Discov Today 12(13/14):540–547

Tabrizi MA, Bornstein GG, Klakamp SL, Drake A, Knight R, Roskos L (2009) Translational strategies for development of monoclonal antibodies from discovery to the clinic. Drug Discov Today 14(5–6):298–305

Terrell TG, Green JD (1994) Issues with biotechnology products in toxicologic pathology. Toxicol Pathol 22:187–193

Thomas JA (1995) Recent developments and perspectives of biotechnology-derived products. Toxicology 105:7–22

Tibbitts J, Cavagnaro JA, Haller CA, Marafino B, Andrews PA, Sullivan JT (2010) Practical approaches to dose selection for first-in-human clinical trials with novel biopharmaceuticals. Regul Toxicol Pharmacol 58:243–251

Vahle JL, Finch GL, Heidel SM, Hovland DN, Ivens I, Parker S, Ponce RA, Sachs C, Steigerwalt R, Short B, Todd MD (2010) Carcinogenicity assessments of biotechnology-derived pharmaceuticals: a review of approved molecules and best practice recommendations. Toxicol Pathol 38:522–553

Vargas HM, Bass AS, Breidenbach A, Feldman HS, Gintant GA, Harmer AR, Heath B, Hoffman P, Lagrutta A, Leishman D, McMahon N, Mittelstadt S, Polonchuk L, Pugsley MK, Satata JJ, Valentin J-P (2008) Scientific review and recommendations on preclinical cardiovascular safety evaluation of biologics. J Pharmacol Toxicol Methods 58:72–76

Vidal JM, Kawabata TT, Thorpe R, Silva-Lima B, Cederbrant K, Poole S, Mueller-Berghaus J, Pallardy M, Van der Laan JW (2010) In vitro cytokine release assays for predicting cytokine release syndrome: the current state-of-the-science. Report of a European Medicines Agency Workshop. Cytokine 51(2):213–215

Weinbauer GF, Niehoff M, Niehaus M, Srivastav S, Fuchs A, Van Esch E, Cline JM (2008) Physiology and endocrinology of the ovarian cycle in macaques. Toxicol Pathol 36(7S):7S–23S

Zbinden G (1987) Biotechnology products intended for human use, toxicological targets and research strategies. In: Grahm CE (ed) Preclinical safety of biotechnology products intended for human use: clinical and biological research, vol 235. Alan R. Liss, New York, pp 143–159

Zbinden G (1990) Safety evaluation of biotechnology products. Drug Saf 5(suppl 1):58–64

Zbinden G (1991) Predictive value of animal studies in toxicology. Regul Toxicol Pharmacol 14:167–177

Chapter 11
Safety Pharmacology: Guidelines S7A and S7B *

John E. Koerner and Peter K.S. Siegl

Abstract Safety pharmacology studies are performed during nonclinical drug development to identify and characterize, in relationship to exposure, potentially undesirable pharmacodynamic effects of a substance on physiological functions. A major objective of these studies is to assess the relevance of these pharmacodynamic activities for human safety. The International Conference on Harmonisation (ICH) issued guidelines describing nonclinical safety pharmacology testing strategies to detect effects on core systems, that is, cardiovascular, respiratory, and central nervous systems (ICH S7A), and risk of delaying ventricular repolarization (QT interval prolongation) (ICH S7B). An ICH Expert Working Group (EWG) took on the task of developing safety pharmacology guidelines and achieved step 4 with ICH S7A in 2001. Drug-induced delay in ventricular repolarization (QT interval prolongation) is the topic of a complementary guideline, ICH S7B, which had many of the same EWG members and achieved step 4 in 2005. The present chapter describes these guidelines along with background and context for the final recommendations in the guidelines.

* This article reflects the personal opinions of the author and does not necessarily reflect those of the FDA.

J.E. Koerner, Ph.D. (✉)
United States Food and Drug Administration, Center for Drug Evaluation and Research, Silver Spring, MD, USA
e-mail: John.Koerner@fda.hhs.gov

P.K.S. Siegl
Siegl Pharma Consulting LLC, Blue Bell, PA, USA

J.W. van der Laan and J.J. DeGeorge (eds.), *Global Approach in Safety Testing*, AAPS Advances in the Pharmaceutical Sciences Series 5, DOI 10.1007/978-1-4614-5950-7_11,

11.1 Introduction

Safety pharmacology originated as a scientific discipline, based on the observation that, in addition to findings in toxicology studies, pharmacodynamic (functional) effects can have clinical safety significance (Bass et al. 2004). These effects may not be readily captured in traditional toxicology studies. It was noted that "The adverse drug reactions [that] the standard toxicological test procedures do not aspire to recognize include most of the functional side effects. Clinical experience indicates, however, that these are much more frequent than the toxic reaction due to morphological and biochemical lesions…" (Zbinden 1979). Additionally, the origin of safety pharmacology guidelines was recently described by Pugsley et al. (2008). Regulatory authorities and sponsors had a common interest in being able to capture pharmacodynamic effects in nonclinical studies that are not captured in traditional toxicology studies.

The ICH safety pharmacology guidelines, S7A "Safety Pharmacology Studies for Human Pharmaceuticals" and S7B "Nonclinical Evaluation of the Potential for Delayed Ventricular Repolarization (QT Interval Prolongation) by Human Pharmaceuticals," describe nonclinical testing strategies to detect and characterize pharmacological activities of potential drug candidates that could impact clinical safety.[1] Since biological systems are complex, the ICH S7A and S7B guidelines emphasize testing for pharmacodynamic (functional) activities of drug candidates using in vivo testing models in which indices of vital organ function are evaluated. Although drug candidates are typically optimized for their therapeutic potential via high potency and selectivity at the therapeutic target, the drug candidate may have additional functional pharmacological activities not revealed in the lead optimization process. The types of pharmacological activities detected in safety pharmacology assays are also not typically evaluated in routine toxicology studies but do have direct corollaries to safety endpoints monitored in clinical studies.

The first reference to safety pharmacology studies in ICH guidelines was in ICH M3, "Guidance on Nonclinical Safety Studies for the Conduct of Human Clinical Trials and Marketing Authorization for Pharmaceuticals." An ICH Expert Working Group (EWG) took on the task of developing a safety pharmacology guideline and achieved step 4 with ICH S7A in 2001. The topic of assessing drug-induced delay in ventricular repolarization (QT interval prolongation) is the topic of a complimentary guideline, ICH S7B, which had many of the same EWG members and reached step 4 in 2005. It has been 6–10 years since the safety pharmacology guidelines were adopted by all three regions.

Developing these two guidelines was controversial because, with the exception of the Japanese Guidelines for General Pharmacology Studies (Ministry of Health, Labour and Welfare of Japan [MHLW]) (Anon 1995), safety pharmacology (also referred to as general pharmacology or ancillary pharmacology) was a function already performed by many sponsors to reduce risk of attrition in drug development.

[1] FDA refers to these guidelines as guidances in accordance with FDA's good guidance practice (62 FR 8961, February 27, 1997).

The timing, design, and types of studies varied among sponsors, reflecting different philosophies and risk tolerance (Bass et al. 2004). In some cases, sponsors considered their approach to be a competitive advantage. The goal of the ICH EWG was to create guidelines that provided direction for sponsors but also maintained flexibility.

11.2 Objectives and General Principles of Safety Pharmacology in Drug Discovery and Development Programs

A primary goal of safety pharmacology studies is to protect clinical trial participants and patients. Information from these studies can also aid in selection of the clinical candidates, doses, and design of clinical programs as well as reduce risk of attrition due to drug-related adverse effects during all phases of development. To accomplish this most effectively and to minimize use of resources and animals, the safety pharmacology guidelines recommend a scientific and efficient approach in the choice and design of assays, as well as interpretation of results (see Table 11.1). Safety pharmacology studies are usually performed during characterization of a development candidate and prior to initiation of clinical studies. At this stage there are data on selectivity from in vitro screens (receptors, enzymes, and ion channels), metabolism, and characterization of the targeted pharmacological activity.

11.2.1 Safety Pharmacology Assays

Sponsors are encouraged to consider assays to evaluate other organ systems and/or endpoints based upon knowledge of selectivity profile and chemical/pharmacological class of the development candidate. In each guideline there is a recommended core battery of assays. These core assays (cardiovascular, central nervous, and respiratory systems) were considered to be vital, since adverse effects on their functions can be acutely life threatening. Results from core assays provide a standard set of data and are expected to be included in regulatory documents, unless there is justification for not doing these assays. In addition to data from the core battery, the sponsor should consider whether further characterization of the activity in other organ systems or with follow-up studies will provide more complete information for a better risk assessment. In this way, the EWG encouraged sponsors to gather information that can most effectively characterize the safety of the drug candidate.

Both guidelines encourage use of conscious, unrestrained animals based upon the premise that autonomic reflexes are intact and this setting is therefore "more physiological." Also use of conscious animals is analogous to assessing responses in conscious human subjects. The EWG understood that there could be direct pharmacological activity that is more easily detected in anesthetized preparations or

Table 11.1 Steps in scientific approach to profiling drug candidates with safety pharmacology studies

1. Identify mechanism and non-mechanism-based pharmacological activities of drug candidates in major organ systems using functional endpoints (in vivo).
2. Characterize these pharmacological activities: a. Relative potency—dose and concentration relative to human. b. Mechanism of action.
3. Compare activity and potency to reference drugs with clinical experience.
4. Estimate relative risk of these activities for potential adverse effects in humans using all available data (safety pharmacology, toxicology, metabolism and clinical experience, target patient population, and concomitant medications).

in vitro models and, consequently, one should not dismiss results from in vitro or anesthetized preparations when they do not appear to be consistent with results in conscious preparations. The guidance allowed for safety pharmacology endpoints to be captured in toxicology studies, which also utilize conscious animals—with the caveat that these core parameters need to be captured in a sensitive manner.

As in general toxicology studies, normal, healthy animals are used in safety pharmacology studies, because this provides the most consistent background for detecting, characterizing, and comparing pharmacological activities. In some cases, animal models of disease may be utilized as follow-up assays to aid in the overall risk assessment. This option was not specifically discussed in the guidelines due to concerns of applicability of animal disease models to broad patient populations. In part, the appropriateness of the specific disease model needs to be justified—and its ability (sensitivity) to capture clinically relevant effects needs to be demonstrated. Use of animal models of disease should be accompanied by data in healthy animals.

An option to conduct safety pharmacology assessments as part of a toxicology study is offered in both ICH S7A and ICH S7B, with the recommendation that assay sensitivity, validation, and quality of data be satisfactory in the toxicology study. The application of this option to assess risk of delayed ventricular repolarization (QT interval prolongation) is discussed below (Section “In Vivo QT Assay”).

11.2.2 Timing of Safety Pharmacology Studies

The safety pharmacology guidelines recommend that the sponsor evaluate test compounds in the core batteries prior to initiation of clinical studies to provide support for first-in-human trials. Sponsors can perform additional nonclinical safety pharmacology studies later during development, for example, to help characterize unanticipated activity observed in toxicology or clinical studies. The guideline encourages sponsors to integrate safety pharmacology findings with those from toxicology, pharmacokinetic, and nonclinical and clinical pharmacology studies for interpretation of overall safety and risk assessments. It is also

important to reexamine safety pharmacology data as well as to consider performing additional studies during development as nonclinical and clinical data accumulate.

11.2.3 Frequency of Dosing, Route of Administration, and Dose Levels

The guidelines recommend acute studies, with a single administration of the test substance via the intended clinical route in healthy animals. The guidelines also recommend that the sponsor determine the time course and dose–response relationship of drug-related effects. These recommendations are consistent with the primary objective of the studies, which is to determine pharmacodynamic (functional) effects of the test substance. The safety pharmacology studies are not intended to mimic the clinical situation or to evaluate how the response is modified by disease. Like toxicology studies, these experimental conditions were chosen as the most consistent setting to detect and characterize pharmacological activities. Implications/consequences of the safety pharmacology actions in humans with or without disease should be considered in the risk assessment (see Table 11.1).

The recommendation for dose levels in safety pharmacology studies parallels the approach in toxicology studies. That is, drugs will have toxicity or unanticipated pharmacological activity at some dose level. The goal of these studies is to identify and characterize dose-limiting pharmacodynamic effects and to determine safety margins to guide clinical testing. Although not required, information on the mechanism(s) of the unanticipated safety pharmacology activity can help in the overall risk assessment by enabling comparisons with drugs sharing the same mechanism(s). Because safety margins can change as additional information becomes available, that is, when therapeutic doses are refined with clinical data, margins should be reevaluated when appropriate.

11.2.4 Assay Sensitivity and Use of Reference Compounds

An important scientific aspect of recommendations in the safety pharmacology guidelines is to interpret and communicate the results in the context of the assay used. Both guidelines encourage reporting data in reference to assay sensitivity and responses to positive and negative controls (see box below).

Reporting that a test compound has no activity in an assay without knowledge of the assay's sensitivity and whether positive controls can be detected is a poor use of resources, and the conclusion can be misleading. It is incorrect to assume that assays employing similar protocols will perform exactly the same in every laboratory, even if performed under Good Laboratory Practices (GLP). To be able to conclude that there is no activity with the test compound in an assay that has a, say, 90 % power to

Text from Safety Pharmacology Guidelines on Assay Sensitivity and Use of Positive and Negative Controls

ICH S7A

"Appropriate negative and positive control groups should be included in the experimental design. In well-characterized in vivo test systems, positive controls may not be necessary. The exclusion of controls from studies should be justified."

ICH S7B

"A sub-maximally effective concentration of a positive control substance should be used to demonstrate the responsiveness of in vitro preparations for ion channel and action potential duration assays and should be included in every study. In the case of in vivo studies, positive control substances should be used to validate and define the sensitivity of the test system, but need not be included in every study."

detect a change of 10 % magnitude and where a clinically relevant positive control can be detected in the assay at a relevant exposure is much more useful than merely concluding that the test compound had no activity in the assay. The EWG recognized that compounds have additional activities at some level; therefore, the goal is identify activity and to report the conditions at which no activity was observed. A common error is to evaluate high doses or concentrations of the positive control (e.g., high dose of dofetilide for QT interval prolongation or I_{Kr}/hERG inhibition); assessment of excessive doses does not adequately assess assay sensitivity. There was much discussion about the need for positive controls, and a compromise was reached based on scientific need, practicality, and animal usage. In general, concurrent positive controls were recommended for in vitro studies, whereas for in vivo studies, it is reasonable to rely on historical control data for that laboratory.

11.2.5 Relationship Between Pharmacodynamic and Pharmacokinetic Data

To satisfy the recommendation in the guidelines that exposure of drug and metabolites will include and exceed targeted exposure in humans, plasma levels of drug and metabolites need to be documented for the dose levels tested. While it is ideal to measure pharmacodynamic (PK) and pharmacokinetic (PD) in same animals to minimize variability, this may not be practical; therefore, PK data from other studies are sometimes used. Note that the use of extrapolated PK values to document exposure can be misleading and result in erroneous estimates of safety margins.

Using such data to support results from safety pharmacology studies is inconsistent with the recommendations in the guidelines.

Because pharmacokinetics and metabolism can differ among species, it is prudent to examine the magnitude and timing of PD effects in relation to plasma levels of parent and, if appropriate, metabolites. This is consistent with interrogating direct pharmacological activities and determination of safety margins. When there is a direct correspondence between time courses of activity and plasma levels, as well as a concentration/dose and magnitude of effect, it strengthens the conclusion that the observed effect is test article related. Reporting activity in terms of plasma levels also facilitates translation of relative potencies (e.g., ED_{50} or IC_{50} concentrations) and thresholds for activity (e.g., NOEL or NOAEL) among species, including humans.

11.2.6 Safety Pharmacology Studies with Biologics

In ICH S7A, "For biotechnology-derived products that achieve highly specific receptor targeting, it is often sufficient to evaluate safety pharmacology endpoints as a part of toxicology and/or pharmacodynamic studies; therefore, safety pharmacology studies can be reduced or eliminated for these products." This is consistent with guidance provided in ICH S6 (Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals): "It is important to investigate the potential for undesirable pharmacological activity in appropriate animal models and, where necessary, to incorporate particular monitoring for these activities in the toxicity studies and/or clinical studies. … These functional indices may be investigated in separate studies or incorporated in the design of toxicity studies."

11.2.7 Good Laboratory Practice

Both guidelines point out the importance of ensuring the reliability and quality of the nonclinical safety pharmacology studies because the data are used to support clinical safety. It is noted that "this is normally accomplished through conduct of studies in compliance with good laboratory practice (GLP)." There are situations, however, where the one or more of the core battery safety pharmacology studies might have been performed at a development stage before it was practical to satisfy all aspects of GLP. The guideline indicated that "data quality and integrity in safety pharmacology studies should be ensured even in the absence of formal adherence to the principles of GLP. When studies are not conducted in compliance with GLP, study reconstruction should be ensured through adequate documentation of study conduct and archiving of data."

11.3 S7A Guideline "Safety Pharmacology Studies for Human Pharmaceuticals"

In 1991, Japan's Ministry of Health and Welfare (MHLW) issued a Guideline for General Pharmacology. This guideline recommended evaluation of drug candidates in a panel of in vitro and in vivo assays to assess direct pharmacological activity of drug candidates on many vital functions including the autonomic nervous system (Anon 1995). There were no similar guidelines issued by regulatory agents in other regions. Most sponsors appreciated the value of general or safety pharmacology studies to support selectivity of drug candidates and design of early clinical trials; however, the strategy varied among sponsors based upon their experience and risk tolerance (Bass et al. 2004). The goal of the ICH S7 EWG was to develop a guideline that provided practical direction consistent with the objectives in ICH M3 for studies recommended to be performed prior to initiating clinical studies. For sponsors already doing these type studies, the core battery was usually a portion of their packages, and, for those sponsors who were not doing these studies, the guideline was to facilitate their accomplishing this goal. A key recommendation from the EWG was that the core safety pharmacology studies should be performed before initiating clinical studies to aid in both design of and interpretation of results from the clinical development program. A key goal was to provide for an additional measure of safety in the first-in-human study.

The EWG did not include the in vitro studies from the MHLW guideline in S7A guidance because they wanted to provide flexibility in how sponsors handled in vitro selectivity screening. The ICH S7A guideline refers to in vitro data and recommends using results from the in vitro studies to select and design the safety pharmacology studies. By choosing to focus on evaluation of functional endpoints for vital organ systems, the results from the safety pharmacology studies should reflect the consequences of off-target activities that are elucidated in the in vitro screens. Also, the in vitro data can be invaluable in interpreting findings from the in vivo studies.

Dose selection for in vivo studies was somewhat controversial and engendered extensive discussion by the EWG. Because the purpose of these studies is to capture clinically relevant findings, therapeutic and supratherapeutic doses were considered to be necessary for inclusion. Acute toxicology studies served to guide dose selection for the safety pharmacology studies. Indeed, the final guideline incorporated the following change from step 2: "The guidance recommends that in the absence of a safety pharmacology response, the highest dose tested should be a dose associated with moderate toxicity. The guidance recommended that the highest dose tested should equal or exceed those doses producing some adverse effects."

The potential for significant adverse effects on various major organ systems was discussed by the EWG. It was agreed that adverse effects on cardiovascular, respiratory, and central nervous system (CNS) carried the greatest risk for catastrophic safety consequences and therefore are included in the core battery. Evaluation of safety pharmacology effects on other organ systems such as renal,

gastrointestinal, and autonomic nervous system is described under supplemental safety pharmacology studies with the recommendation that evaluating these organ systems should be considered when there is a cause for concern. Some sponsors routinely evaluate all of these organ systems in their safety pharmacology package, despite this not being a recommendation in the guideline. The guideline also suggests consideration of whether there is sufficient information available from toxicology studies to support safety in humans. The EWG expects that the sponsor will design the safety pharmacology evaluations in view of all of the information available for the test compound.

11.3.1 Central Nervous System

The functional observation battery (FOB) (Mattsson et al. 1996) and modified Irwin's Test (Irwin 1968) in mice have a long history of use in evaluating safety of chemicals. The EWG determined that these assays are appropriate for detecting significant, pharmacologically mediated changes in motor activity, behavior, coordination, sensory/motor reflex responses, and body temperature in a standard, straightforward manner with a minimum of resources. This assay is included in the MHLW General Pharmacology Guideline, and the history of use of these assays in the chemistry industry provides a comforting database. By performing the assay under GLP, it is expected that persons conducting the assay be adequately trained and results with test substances are compared to positive and negative controls. Examples of more detailed CNS evaluations are mentioned in the ICH S7A guideline as follow-up assays. Drug dependence liability assessments are sometimes considered in the safety pharmacology scope, but are not discussed in ICH S7A because they are not acutely life threatening and therefore not needed to support the early clinical studies.

11.3.2 Cardiovascular System

Adverse effects on the cardiovascular system are one of the most common reasons for discontinuation of development of promising drug candidates (Laverty et al. 2011) and have potential for severe adverse consequences. The EWG recommended that changes in heart rate, blood pressure, and electrocardiogram (ECG) be evaluated in the core battery cardiovascular assay and other indices, such as cardiac output, cardiac contractility, and peripheral vascular resistance, be considered in the follow-up assays. The reasoning was that significant (major) changes in cardiac and vascular function would be reflected in the endpoints in the core battery. It was recognized that there could be small effects on cardiac function or vascular resistance that will not be reflected in blood pressure and heart rate signals. However, if the magnitude of change in these parameters at dose levels many multiples over therapeutic levels is small, the safety risk in the clinic will be minimal. It is also recognized that functional cardiovascular effects are routinely and easily assessed in the clinical safety studies (phase 1).

At the time that ICH S7A was being discussed, no scientific consensus existed on the preferred approach to addressing risks for repolarization-associated ventricular tachyarrhythmia (i.e., Torsade de Pointes). Additionally there was no internationally recognized guidance on this topic. The EWG determined that this topic would be best served by a separate guideline that could bring the latest evolving information together (see Sect. 11.4).

11.3.3 Respiratory System

Respiratory distress and acute bronchoconstriction are major clinical adverse events with potentially life-threatening consequences. Prior to ICH S7A, respiratory function was generally assessed in nonclinical pharmacology and toxicology studies via observation of depth and pattern of breathing. The EWG initially concluded that this was adequate; however, a case was made for more quantitative indices of respiratory function to support the safety of new drug candidates. The EWG concluded that more quantitative indices of respiratory function better supported the safety of new drug candidates. Therefore, the following change was incorporated following step 2: "The guidance recommends that, in addition to respiratory rate, other measures of respiratory function (e.g., tidal volume or hemoglobin oxygen saturation) should be evaluated in assessing effects of the test substance on the respiratory system."

11.3.4 Supplemental Safety Pharmacology Studies

Studies to evaluate safety pharmacology effects in other organ systems are listed as supplemental studies. This is included in the ICH S7A guideline for sponsors to consider whenever there are potential safety concerns in other organ systems that are not evaluated in the core battery or other toxicology studies. As mentioned above, many sponsors have included other organ systems in their safety pharmacology packages.

11.4 S7B Guideline "Nonclinical Evaluation of the Potential for Delayed Ventricular Repolarization (QT Interval Prolongation) by Human Pharmaceuticals"

11.4.1 Background and Objectives

Several drugs were removed from the market when it was recognized that they were associated with deaths due to a ventricular tachycardia called Torsade de Pointes (TdP). All of these drugs delayed ventricular repolarization (prolonged the QT interval of the surface ECG) via inhibition of a delayed rectifier potassium channel, I_{Kr} (Darpo 2001; Redfern et al. 2003). I_{Kr} is commonly referred to as the hERG channel.

The human Ether-à-go-go-Related Gene (hERG) is responsible for expression of the I_{Kr} channel protein, and hERG is used to express the human protein in cell lines for I_{Kr} bioassays. Identification of a molecular mechanism (I_{Kr} inhibition) that contributes to risk of TdP and availability of an accessible index of delayed ventricular repolarization (QT interval prolongation on the surface ECG) in animals and humans provides opportunity to evaluate the risk for this type of cardiac activity during drug development.

In 1997, "Points to Consider: The Assessment of the Potential for QT Interval Prolongation by Non-Cardiovascular Medicinal Products" (CPMP/986/96) was issued by the Committee for Proprietary Medicinal Products (CPMP) (Anon 1997). This was the first regulatory document to describe a nonclinical testing strategy for assessing risk of QT interval prolongation as a means to reduce the risk of drug-induced TdP. The document recommended measurement of action potential duration (APD) in in vitro cardiac preparations (e.g., rabbit Purkinje fiber) and changes in ECG QT interval duration in in vivo animal models. While the scientific rationale for employing these two assays to assess risk for a drug to prolong ventricular repolarization (increases in APD at the cardiac cellular level and QT interval at the surface ECG) was sound, there were questions about the reliability of these assays and how to use these data for risk assessments in humans. Also, at the time of these recommendations, the role of I_{Kr} inhibition as a common molecular mechanism for drug-induced prolongation of ventricular repolarization was not known.

With this background, the task for the ICH S7B EWG was to develop a guideline using the CPMP document, as well as a draft guidance from Health Canada (Strnadova 2005) as starting points. Outstanding issues to be addressed by the EWG included translation of I_{Kr} inhibitory potencies to risk of QT interval prolongation, accurate measurement of QT interval duration as a reliable index of changes in ventricular repolarization, relationship between QT interval prolongation, and TdP. It was first determined that a guideline could provide value by recommending a testing strategy to assess the risk of delayed ventricular repolarization (QT interval prolongation), but it was unrealistic at that time to develop guidelines for assessing the risk for drug-induced TdP arrhythmia. Therefore, the title and objective of ICH S7B refer to assessing the risk of delayed ventricular repolarization and not the proarrhythmia risk for drug candidates.

It is important to note that the safety concern was an unexpected cardiac toxicity associated with mortality, which had occurred with several noncardiac drugs. While the incidence of toxicity was low, some of the drugs such as the antihistamine, terfenadine (Seldane®), were widely prescribed; therefore, the risk was considered unacceptable when considered over the population of users at large. Recognition of the relationship between delayed ventricular repolarization and risk for TdP was confirmed from investigation of the genetic QT prolongation syndrome as well as drug-induced QT interval prolongation. In both of these scenarios, the QT interval prolongation was only one of several risk factors that needed to be present at the same time to induce the arrhythmia and hence the very low incidence of arrhythmias even when the QT interval is delayed. Again, the objective of S7B is to assess the risk for delayed ventricular repolarization by drugs as a strategy to reduce one of the risk factors for TdP. Because of the low incidence of TdP, determining if the strategy

reduces the risk of TdP requires very large patient experience (i.e., absence of TdP in clinical trials prior to registration is usually not sufficient to exclude this risk).

After the ICH S7B EWG was underway, developing a guideline for the clinical assessment risk of QT interval prolongation becomes an ICH topic. The ICH E14 EWG began working on their guidance, "Clinical Evaluation of QT/QTc Interval Prolongation and Proarrhythmic Potential for Non-Antiarrhythmic Drugs." During this time, there were joint EWG meetings so that the two guidelines could be aligned, and it was agreed to publish both at the same time. A controversial issue was the need for clinical assessment of QT interval prolongation risk when the nonclinical assessment indicated there was very low risk. At the time of introduction of the guidelines, there was no prospective experience to conclude the clinical studies were not needed; however, this has been an ongoing discussion and a topic of great interest to all involved (Trepakova et al. 2009).

11.4.2 Nonclinical Studies Performed to Support ICH S7B

The interest in achieving a practical and effective guideline was shared by the pharmaceutical industry, academia, and regulatory agencies. As a result there were several studies performed (and published) that provided useful information for the ICH S7B EWG. One was an investigation performed under the guidance of the ILSI-HESI Cardiovascular Safety Subcommittee where positive and negative control drugs (all with clinical experience) were prospectively tested in three nonclinical assays: inhibition of I_{Kr} in vitro, APD prolongation in vitro (canine Purkinje fiber), and QT interval prolongation in vivo (conscious dogs instrumented with telemetry). The results (Hanson et al. 2006) demonstrated that (1) the I_{Kr} inhibitory potency was reliably measured in two independent laboratories using cells lines with hERG expression of I_{Kr}, (2) the canine Purkinje fiber APD assay had a significant number of false negatives, and (3) the in vivo assay correctly identified all drugs with QT interval prolonging activity. Members of Japanese Pharmaceutical Manufacturers Association (JPMA) also did a series of prospective studies (Ando et al. 2005; Hayashi et al. 2005; Kii et al. 2005; Miyazaki et al. 2005; Omata et al. 2005; Sasaki et al. 2005; Tashibu et al. 2005; Toyoshima et al. 2005; Yamazaki et al. 2005) expanding the findings of the ILSI-HESI group to the guinea pig papillary muscle for APD assessment and nonhuman primate for in vivo QT assessment. The Association of the British Pharmaceutical Industry (ABPI) group lead by Tim Hammond performed a retrospective study to determine acceptable safety margins for potency at the I_{Kr} channel (Redfern et al. 2003). Their conclusion was that when the margins (adjusted for plasma protein binding) are greater than 30-fold, the risk for QT interval prolongation is low. This of course is a very broad generalization, but does support the concept that margins should be considered and not all inhibitors of I_{Kr} have the same risk for adverse CV effects. The ICH S7B EWG was fortunate to have these prospective and retrospective data, as well as scientific input from scientists in all three ICH regions.

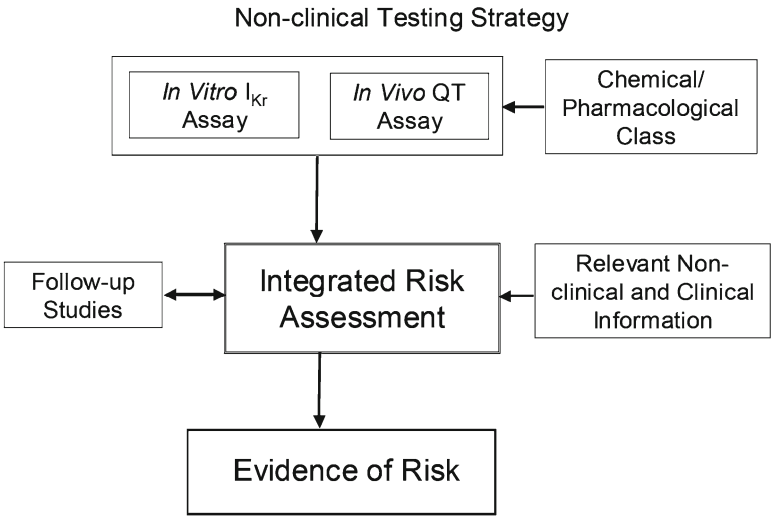


Fig. 11.1 Nonclinical testing strategy from ICH S7B guideline

11.4.3 Testing Strategy and Assay Selection

The ICH S7B EWG created Fig. 11.1 to illustrate the general testing strategy. The EWG recommends that the sponsor consider whether the test substance belongs to a pharmacological or chemical class that is associated with a known risk for QT interval prolongation and/or TdP. For example, many antipsychotic and non-sedating antihistamine drugs have been associated with QT interval prolongation and TdP in humans. In such cases, the sponsor is encouraged to pursue testing that strategy that directly compares the test compound to those in the same class with documented risk.

Following the precedent to recommend a combination of core battery, follow-up, and supplementary assays in ICH S7A (see above), the ICH S7B EWG spent significant time debating the assays that should be the core battery assays.

Because all of the drugs removed from the market due to an association with TdP delayed ventricular repolarization by inhibiting I_{Kr} and because the human form of the channel protein can be expressed in cell lines, an In Vitro Ion Channel Assay was included in the core battery (see section "In Vitro Ion Channel Assay").

Testing for the potential for drug candidates to prolong cardiac APD is recommended in the CPMP Points to Consider document (see Sect. 11.4.1) and is a logical step to determine if inhibition of I_{Kr} detected in the In Vitro Ion Channel Assay translates into APD prolongation in a multicellular preparation. However, based upon EWG experience, as well as the results from the ILSI-HESI study (Hanson et al. 2006), there was concern about the high incidence of false-negative results in the Purkinje fiber APD assay. It was recognized that when activity is detected in in vitro APD assays, the results can be very important in characterizing the relative

risk and potential effects of the test compound on other APD parameters, including other cardiac ion channels. As a result, the APD assay was included as a follow-up assay (see section "Follow Up Assays"). Other in vitro assays such as the rabbit Langendorff heart (Hondeghem et al. 2001; Hondeghem 2006) and the ventricular wedge preparation (Yan and Antzelevitch 1996; Liu et al. 2006) that measure additional characteristics of repolarization (e.g., instability of APD changes and dispersion of refractoriness, respectively) were discussed and included in the supplementary assays due to their technical complexity and their focus on arrhythmia risk rather than simply duration of ventricular repolarization.

An in vivo assay directly measuring QT interval duration was included in the core battery because it has the potential to detect effects of drug candidates on ventricular repolarization by any mechanism or combination of mechanisms. As shown in Fig. 11.1, the In Vivo QT Assay is the final step in the core battery because it integrates a drug's effects on ventricular repolarization, and is analogous to the clinical setting (including the bioassay recommended in ICH E14) to assess risk of QT interval prolongation in humans (see section "In Vivo QT Assay").

11.4.3.1 In Vitro Ion Channel Assay

All drugs that have a direct inhibitory effect on I_{Kr} will delay ventricular repolarization in vivo when appropriate plasma concentrations are achieved in the heart, and there are no other electrophysiological effects that modulate the effects of I_{Kr} on ventricular repolarization. The basic pharmacological principle is that inhibition with selective I_{Kr} blockers is concentration related and relative potency data can be used to compare compounds and estimate safety margins, as was shown by Redfern et al. (2003). The In Vitro Ion Channel Assay uses the human protein; however, because the structure and therefore pharmacology of I_{Kr} are similar across species, translation of relative potencies at the I_{Kr} channel level—from in vitro human to in vivo dog, nonhuman primate or swine—is very good, and translation is acceptable from nonclinical in vivo to clinical settings.

It is important to recognize that translation of in vitro potency into in vivo activity is influenced by factors that affect access of the test compound to the I_{Kr} ion channel, such as metabolism, distribution, and plasma protein binding. Also, when the test compound has effects on multiple cardiac ion channels, estimating safety margins from in vitro I_{Kr} inhibitory potencies alone is difficult. Therefore, not all I_{Kr} blockers will prolong the QT interval in vivo at exposure levels where in vitro activity was observed. Note that ICH S7B does not make recommendations about the appropriate safety margin for the test substance because factors such as therapeutic indication (benefit–risk assessment), disposition, and other pharmacological characteristics (safety margin) should be considered by the sponsor. Also because of the complexity in predicting relative potencies in vivo, the in vitro potency values and safety margins are refined when in vivo data are available.

At the time of development of ICH S7B, the in vitro assay for assessing relative potency of I_{Kr} inhibition used standard voltage clamp methodology. This is a

technically challenging assay. The EWG recognized that ligand binding assays were available, but, based upon the low specific activity of most the available radioligands, results are generally not robust enough for risk assessment. Since publication of ICH S7B, high-throughput voltage clamp assays have become available and can be adequate for the In Vitro Ion Channel Assay if sensitivity and specificity are defined. A practical problem with some of these new assay systems is binding of lipophilic compounds to the plastic in the high-throughput instruments, which may underestimate the inhibitory potency.

From clinical experience with drugs and information from congenital long QT syndrome, the ICH S7B EWG was aware that there are cardiac ion channel mechanisms in addition to I_{Kr} inhibition that can delay ventricular repolarization in a manner that are risk factors for TdP. These include inhibition of I_{Ks}, agonism of the window sodium channel, and modulation of cardiac calcium channels. There is clear value in assessing the relative potencies of test substance on these other mechanisms early in the evaluation process (Hancox et al. 2008); however, the EWG concluded, because of the promiscuous behavior of the I_{Kr} channel for inhibition by drugs (Sanguinetti and Mitcheson 2005; Sanguinetti and Tristani-Firouzi 2006), that this was the mechanism of greatest risk. It was also reasoned that if the other, less commonly seen mechanism(s) were present, they would be detected in the In Vivo QT Assay. In fact, if QT interval prolongation is observed in the In Vivo QT Assay that is inconsistent with the test substances inhibitory potency on I_{Kr}, the use of follow-up assays to explore the effects on other ion channels is prudent. Therefore, the decision to screen for mechanisms in addition to I_{Kr} inhibition is left up to the sponsor, dependent on their risk tolerance for a possible non-I_{Kr} mechanism QT interval prolongation appearing in the In Vivo QT Assay.

Since the publishing of ICH S7B, there is evidence that drugs can interfere with the "trafficking" of the I_{Kr} channel protein to the surface of the cell. This is a potential mechanism for drugs to prolong the QT interval without directly inhibiting the I_{Kr} channel (Delisle et al. 2004; Hancox and Mitcheson 2006). The turnover rates of the I_{Kr} proteins or pharmacodynamic relationship between level of inhibition and delay in repolarization are not known, making it challenging to interpret the relative risk of QT interval prolongation from the available in vitro trafficking assays. The ICH S7B EWG did not discuss this topic, so this mechanism is not included in the guideline.

11.4.3.2 In Vivo QT Assay

QT interval prolongation of the electrocardiogram (ECG) is a consequence of APD prolongation at the cellular level and delayed ventricular repolarization at the organ level. Therefore, measuring the QT interval duration in relevant animal models and in humans is a practical approach to assessing delay in ventricular repolarization. As such, the In Vivo QT Assay is a central component of the S7B testing strategy and relates directly to objectives and endpoints in ICH E14 and other clinical safety testing.

The cardiovascular assays in the ICH S7A core battery and the GLP toxicity studies include evaluation of the ECG; however, it was recognized that additional considerations are required to evaluate the risk of drug-induced delayed ventricular repolarization. This was an important topic for the ICH S7B EWG in developing the guidance. First, the species used to assess risk of QT interval prolongation in humans needs to be considered. Unlike humans, the duration of ventricular repolarization is not controlled by I_{Kr} in rodents, and therefore, one cannot assess the risk of QT interval prolongation for humans using rats or mice. In the ICH S7A guidance, there is no recommendation for species in the cardiovascular assessment (see Sect. 11.3.2). If the sponsor chooses to use rodents for the ICH S7A assessment, an additional study in non-rodents is needed to investigate the effects on ventricular repolarization and comply with recommendations in ICH S7B. Second, the sensitivity of ECG recordings to detect changes in QT interval is rather poor in toxicology studies due to high sympathetic tone and variable heart rates with methods of restraint and brief sampling periods. The availability of implantable telemetry devices for dogs and nonhuman primates as well as computer-assessed measurement of ECG intervals provided an opportunity to capture high-quality ECG signals and evaluate many complexes over a long period. The ICH S7B guideline does not specifically recommend the use of telemetry but does recommend determining sensitivity and specificity of the assay/method used to support the risk assessment. Since finalization of ICH S7B, there are now alternatives such as jackets that can capture ECG data with reasonable quality without surgical implantation of a device (Chui et al. 2009; Kyle et al. 2009). Both ICH S7A and S7B describe an option to collect ECG data for QT intervals in the toxicology studies with the premise that sensitivity and specificity need to be defined in order to support conclusions from the data. Note that the level of sensitivity for detecting QT intervals is not dictated in the guideline, but the suggestion is that the sponsors use an assay that has sensitivity appropriate for the risk. It has been challenged whether toxicology studies can adequately assess risk of drug-induced QT interval changes; however, in principle, QT interval data from toxicology studies will be in compliance with ICH S7B if guideline recommendations concerning sensitivity are satisfied (see Guth et al. 2009).

Measurement of changes in QT interval duration as an index of ventricular repolarization is not straightforward. The duration of the QT interval is significantly affected by changes in the heart rate, respiratory patterns, and autonomic nervous system activity. The ICH S7B EWG discussed the value of assessing changes in QT interval in anesthetized preparations where some of these variables can be controlled; however, the consensus was that the conscious, unrestrained animal would be the more appropriate setting for predicting risk in humans. It is recommended that sponsor consider using the anesthetized preparation when there are drug-induced changes in sympathetic tone or as a follow-up assay to determine if changes detected in the conscious preparation are direct effects on ventricular repolarization or a consequence of altered autonomic tone and/or overcorrection with QT interval heart rate correction formulae.

There is no absolutely reliable method to adjust QT interval duration measurements for changes in heart rate or autonomic tone. There are several correction formulae (Miyazaki and Tagawa 2002) which are typically valid over small changes in heart rate. Given this dilemma, ICH S7B makes no recommendations beyond justifying the choice of heart rate correction formula with data from the test system. The guideline

also recommends that sponsors consider analyzing the data by plotting the QT/RR relationship. Heart rate correction of QT intervals is also an issue for the corresponding clinical assay in ICH E14. In many cases, sponsors will analyze the data with several formulae and discuss the totality of data set to support their conclusions.

Species differences in potencies for QT interval prolongation in vivo are not due to species differences at the channel level (i.e., relative potency for I_{Kr} inhibition; see section "In Vitro Ion Channel Assay"). They can be due to differences in distribution, metabolism, plasma protein binding, background autonomic tone (including baseline heart rate), and other cardiovascular effects. Therefore, the guideline makes no specific recommendations as to a preferred species for this assay but does recommend that the sponsor select and justify the most appropriate in vivo test systems and species.

11.4.3.3 Follow-Up Assays

As discussed above (Sect. 11.2.1), the objective of follow-up assays is to obtain additional information to interpret and/or provide context for results from assays in the safety pharmacology core battery, pharmacology and toxicology studies, and clinical studies.

For example, when results from the In Vitro Ion Channel and In Vivo QT Assays are not consistent with one another, there are several options for follow-up assays. When there is in vivo but not in vitro activity, testing metabolites for I_{Kr} inhibitory potencies is prudent. If the test compound is active in an in vitro APD assay (Purkinje fiber assay, Langendorff heart preparation, or ventricular wedge assay), assessment of configuration of APD prolongation (APD 30 vs. APD 90) can be helpful in evaluating the consequences of multiple ion channel activities. As mentioned above (Sect. 11.4.1), if the test compound does not prolong the APD (but does prolong QT interval in vivo), results from the APD assay will not be useful. Another follow-up strategy is to test the potencies on other cardiac ion channels. To determine if the heart rate correction formulae might be overcorrecting the duration of the QT interval, a beat-to-beat analysis of the relationship between heart rate and QT duration may be helpful (Fossa et al. 2005).

When the risk of QT interval prolongation is defined for a test compound, follow-up assays are sometimes employed to determine if the proarrhythmic risk is consistent or less than expected from the change in repolarization (see Sect. 11.5).

11.4.3.4 Assays/Strategies for Assessing Proarrhythmia Risk

Assessment of the safety of drug candidates in simulated pathological conditions and arrhythmias is very challenging because of the abundant combinations of risk factors in the broad patient populations. The ICH S7B EWG did not provide specific guidance on this, but provided the following statement in the document: "Interested parties are encouraged to develop these models and test their usefulness in predicting risk in humans."

QT interval prolongation is only one of the several risk factors that must be coincident to be a trigger for TdP (Kowey and Malik 2007); therefore, the incidence of TdP, even

when there is QT interval prolongation, is very small (Darpo 2001, 2007). Because the combinations of risk factors are many and the incidence of TdP is so low, one cannot typically exclude risk of TdP from data in a typical clinical development program. Unless there is a high incidence of TdP, exclusion of risk will usually require post-marketing data. Therefore, when there is a risk for QT interval prolongation at or near therapeutic levels, the label of an approved drug will carry a warning of potential risk for TdP. In this case, prior to approval to market, the sponsor may want to determine how this risk relates to other drugs in the class with and without a significant risk of TdP.

There have been at least two symposia addressing this issue, one by the European Society of Cardiology (Haverkamp et al. 2000) and one by ILSI-HESI (Bass et al. 2004). In both symposia, measurable attributes of test substances that might signal increased risk of TdP were discussed, including dispersion of refractoriness, instability of repolarization, and changes in action potential configuration. No single assay has been proposed, and the prediction is that a battery of nonclinical assays will be needed. There is a case study since launch of ICH S7B where the sponsor successfully made the case for a low risk of TdP despite a clear risk of QT interval prolongation. This case is ranolazine where both nonclinical and clinical data were used. The combination of pharmacological activities of ranolazine on I_{Kr} and I_{Na} was shown to (1) prevent the expected APD prolongation and incidence of early after depolarizations with a potent I_{Kr} blocker, (2) have less than expected transmural dispersion of refractoriness compared to I_{Kr} blocking drugs with a history of TdP in a ventricular wedge preparation (Antzelevitch et al. 2004), and (3) exhibit a decline in incidence of ventricular tachycardia in patients with non-ST segment elevation acute coronary syndrome (Schram et al. 2004; Song et al. 2004). Therefore, demonstrating a lower than expected risk for TdP with a drug that blocks I_{Kr} and prolongs the QT interval prolongation requires a well-designed strategy with use of positive and negative reference agents. There are likely to be regional differences in how regulators interpret these data.

11.4.4 Integrated Risk Assessment and Evidence of Risk

The ICH S7B guideline recommends that an Integrated Risk Assessment for QT interval prolongation be used to maximize the value of the experimental data by considering all of the available information including the targeted indication and patient population. Information on the sensitivities of the nonclinical assay used, relative potencies of the test compound compared to reference drugs, characteristics of the primary pharmacology that could impact risk for QT interval prolongation, and risk of greater exposure due to hepatic impairment or drug–drug interactions are important components. The Integrated Risk Assessment is an important opportunity for the sponsor to make a scientific case that either the risk of QT interval prolongation with their development candidate is negligible at therapeutic levels or the benefit/risk assessment is acceptable for the intended use and indication(s). The objective of generating an Integrated Risk Assessment is to enable prudent decisions by sponsors and regulators, as well as to provide information that can be used to help describe nonclinical data in future labels. The Integrated Risk Assessment should be updated as additional data become available,

including clinical data. The guideline recommends that the Integrated Risk Assessment be included in the Investigator's Brochure and the Nonclinical Overview (ICH M4). Including an Integrated Risk Assessment in regulatory documents is a valuable opportunity for sponsors to insure their data are presented in the most effective way.

The concept of evidence of risk was included in ICH S7B to emphasize that the evaluation of risk is not an all-or-nothing proposition. Initially, it was the intention of the EWG to provide a qualitative scale for ranking and communicating the relative risk for the test compound to prolong the QT interval. This turned out to be too ambitious given the complexity and spectrum of data and indications. Describing evidence of risk in a very qualitative manner in ICH S7B was done to encourage sponsors to provide a context for risk in the Integrated Risk Assessment.

11.4.5 Relationship Between ICH S7B and ICH E14 Guidelines

Ideally, the nonclinical and clinical guidelines should be complementary, and the results from the studies recommended by both are to be used in the risk assessment. The ICH S7B EWG recommended that conclusions from the nonclinical studies would contribute to the design and interpretation of the clinical studies assessing risk of QT interval prolongation. For example, when no risk is identified in the ICH S7B studies, the need for a thorough clinical QT/QTc study would be reduced. Also, "in circumstances where results among nonclinical studies are inconsistent and/or results of clinical studies differ from those for nonclinical studies, retrospective evaluation and follow-up nonclinical studies can be used to understand the basis for the discrepancies" (text from ICH S7B). Analyses of both nonclinical and clinical data would be important to avoid false-positive and false-negative outcomes from either nonclinical or clinical studies. The ICH E14 EWG was not confident that the nonclinical study results would be predictive of the clinical situation and, at the time, there were no prospective data to address this concern. Therefore, at the time of implementation of the guidelines, the results of the clinical assessment alone were deemed the final arbiter of risk for QT interval prolongation in humans. More recently, ILSI-HESI has a project to investigate the concordance among nonclinical and clinical studies as well as the need for a thorough clinical QT/QTc study when no risk is identified in nonclinical studies (Trepakova et al. 2009).

11.5 Post-S7A and Post-S7B Implementation: Lessons Learned and Future Opportunities

Safety pharmacology studies are currently performed by pharmaceutical companies and contract research organizations and have been successfully integrated into preclinical drug development programs (Ewart et al. 2012). Most regulatory filings include data from the assays in the core batteries recommended in ICH S7A and

S7B, and there are usually minimal or no data from the supplementary and follow-up assays. The Safety Pharmacology Society (http://www.safetypharmacology.org) has become a valuable venue for sharing experiences and advancing new ideas and technologies in safety pharmacology (Redfern and Valentin 2011; Cavero 2011).

The recommendations in the guidelines have prompted development of technologies to capture safety pharmacology data such as whole-body plethysmography for assessing indices of respiratory function and implantable/wearable telemetry devices for capturing cardiovascular endpoints in conscious, unrestrained subjects. While this has standardized the assay methodology to a certain degree, there has been a focus on data collection more than interpretation of data and translation to risk.

It would be valuable to evaluate retrospectively the benefit and cost of the ICH S7A- and S7B-recommended studies. Specifically, have these studies effectively reduced attrition of drug candidates? Have they improved the safety of clinical trial participants? Are the resources used to perform these studies in development (i.e., GLP), including use of animals, justified compared to assessing off-target liabilities during lead optimization (see Cavero 2009)?

11.6 Conclusions

The key objectives of the safety pharmacology guidelines are to encourage sponsors to use testing strategies based upon a scientific rationale appropriate for drug candidate, to provide flexibility, and to support interpretation of results in a scientific manner. Such strategies involve validation of assays, definition of sensitivity and specificity, and comparison of results with positive and negative reference drugs (with clinical experience). The recommendations in the guidelines are intended to encourage sponsors to use an evidence-based risk assessment for their compound to support safety for clinical trial participants and patients as well as to reduce attrition of drugs in clinical development.

References

Ando K, Hombo T, Kanno A, Ikeda H, Imaizumi M, Shimizu N, Sakamoto K, Kitani S, Yamamoto Y, Hizume S, Nakai K, Kitayama T, Yamamoto K (2005) QT PRODACT: in vivo QT assay with a conscious monkey for assessment of the potential for drug-induced QT interval prolongation. J Pharmacol Sci 99(5):487–500

Anon (1995) Japanese guidelines for non-clinical studies of drugs manual. Yakuji Nippo, Tokyo

Anon (1997) Committee for Proprietary Medicinal Products (EU). Points to consider: the assessment of the potential for QT interval prolongation by non-cardiovascular medicinal products

Antzelevitch C, Belardinelli L, Zygmunt AC, Burashnikov A, Di Diego JM, Fish JM, Cordeiro JM, Thomas G (2004) Electrophysiological effects of ranolazine, a novel antianginal agent with antiarrhythmic properties. Circulation 110(8):904–910

Bass A, Kinter L, Williams P (2004) Origins, practices and future of safety pharmacology. J Pharmacol Toxicol Methods 49:145–151

Cavero I (2009) Exploratory safety pharmacology: a new safety paradigm to de-risk drug candidates prior to selection for regulatory science investigations. Expert Opin Drug Saf 8(6):627–47

Cavero I (2011) Annual Meeting of the Safety Pharmacology Society: an overview. Expert Opin Drug Saf 2012 11(2):341–353

Chui RW, Fosdick A, Conner R, Jiang J, Bruenner BA, Hugo M, Vargas HM (2009) Assessment of two external telemetry systems (PhysioJacket™ and JET™) in beagle dogs with telemetry implants. J Pharmacol Toxicol Methods 60:58–68

Darpo B (2001) Spectrum of drugs prolonging the QT interval and the incidence of torsades de pointes. Eur Heart J 3(Suppl K):K70–K80

Darpo B (2007) Detection and reporting of drug-induced proarrhythmias: room for improvement. Europace 9(Suppl 4):23–36

Delisle BP, Anson BD, Rajamani S, January CT (2004) Biology of cardiac arrhythmias: ion channel protein trafficking. Circ Res 94:1418–1428

Ewart L, Gallacher DJ, Gintant G, Guillon J-M, Leishman D, Levesque P, McMahon N, Mylecraine L, Sanders M, Suter W, Wallis R, Valentin J-P (2012) How do the top 12 pharmaceutical companies operate safety pharmacology? J Pharmacol Toxicol Methods 66(2):66–70

Fossa AA, Wisialowski T, Magnano A, Wolfgang E, Winslow R, Gorczyca W, Crimin K, Raunig DL (2005) Dynamic beat-to-beat modeling of the QT-RR Interval Relationship: analysis of QT prolongation during alterations of autonomic state versus human ether *a-go-go*-related gene inhibition. J Pharmacol Exp Ther 312:1–11

Guth BD, Bass AS, Briscoe R, Chivers S, Markert M, Siegl PKS, Valentin J-P (2009) Comparison of electrocardiographic analysis for risk of QT interval prolongation using safety pharmacology and toxicological studies. J Pharmacol Toxicol Methods 60:107–16

Hancox JC, Mitcheson JS (2006) Combined hERG channel inhibition and disruption of trafficking in drug-induced long QT syndrome by fluoxetine: a case-study in cardiac safety pharmacology. Br J Pharmacol 149(5):457–459

Hancox JC, McPate MJ, Harchi AE, Zhang YH (2008) The hERG potassium channel and hERG screening for drug-induced torsades de pointes. Pharmacol Ther 119:118–132

Hanson LA, Bass AS, Gintant G, Mittelstadt S, Rampe D, Thomas K (2006) ILSI-HESI cardiovascular safety subcommittee initiative: evaluation of three non-clinical models of QT prolongation. J Pharmacol Toxicol Methods 54:116–129

Haverkamp W, Breithardt G, Camm AJ, Janse MJ, Rosen MR, Antzelevitch C, Escande D, Franz M, Malik M, Moss A, Shah R (2000) The potential for QT prolongation and pro-arrhythmia by non-anti-arrhythmic drugs: clinical and regulatory implications. Report on a Policy Conference of the European Society of Cardiology. Cardiovasc Res 47(2):219–233

Hayashi S, Kii Y, Tabo M, Fukuda H, Itoh T, Shimosato T, Amano H, Saito M, Morimoto H, Yamada K, Kanda A, Ishitsuka T, Yamazaki T, Kiuchi Y, Taniguchi S, Mori T, Shimizu S, Tsurubuchi Y, Yasuda S, Kitani S, Shimada C, Kobayashi K, Komeno M, Kasai C, Hombo T, Yamamoto K (2005) QT PRODACT: a multi-site study of in vitro action potential assays on 21 compounds in isolated guinea-pig papillary muscles. J Pharmacol Sci 99(5):423–37

Hondeghem LM (2006) Thorough QT/QTc not so thorough: removes torsadogenic predictors from the T-wave, incriminates safe drugs, and misses profibrillatory drugs. J Cardiovasc Electrophysiol 17:337–340

Hondeghem LM, Carlsson L, Duker G (2001) Action potential duration prolongation is antiarrhythmic instability and triangulation of the action potential predict serious proarrhythmia, but action potential duration prolongation is antiarrhythmic. Circulation 103:2004–2013

Irwin S (1968) Comprehensive observational assessment: Ia. A systematic, quantitative procedure for assessing the behavioral and physiologic state of the mouse. Psychopharmacologia 13:222–257

Kii Y, Hayashi S, Tabo M, Shimosato T, Fukuda H, Itoh T, Amano H, Saito M, Morimoto H, Yamada K, Kanda A, Ishitsuka T, Yamazaki T, Kiuchi Y, Taniguchi S, Mori T, Shimizu S, Tsurubuchi Y, Yasuda S, Kitani S, Shimada C, Kobayashi K, Komeno M, Kasai C, Hombo T, Yamamoto K (2005) QT PRODACT: evaluation of the potential of compounds to cause QT

interval prolongation by action potential assays using guinea-pig papillary muscles. J Pharmacol Sci 99(5):449–57
Kowey PR, Malik M (2007) The QT interval as it relates to the safety of non-cardiac drugs. Eur Heart J 9(Supplement G):G3–G8
Kyle AM, Rogers PI, Han S, Chen P-S, March KL (2009) LifeShirt® acquisition system to monitor ECG from ambulatory swine and the implementation of an arrhythmia detection algorithm. Conf Proc IEEE Eng Med Biol Soc 2009:4820–4823
Laverty HG, Benson C, Cartwright EJ, Cross MJ, Garland C, Hammond T, Holloway C, McMahon N, Milligan J, Park BK, Pirmohamed M, Pollard C, Radford J, Roome N, Sager P, Singh S, Suter T, Suter W, Trafford A, Volders PGA, Wallis R, Weaver R, York M, Valentin JP (2011) How can we improve our understanding of cardiovascular safety liabilities to develop safer medicines? Br J Pharmacol 163:675–693
Liu T, Brown BS, Wu Y, Antzelevitch C, Kowey PR, Yan G-X (2006) Blinded validation of the isolated arterially perfused rabbit ventricular wedge in preclinical assessment of drug-induced proarrhythmias. Heart Rhythm 3:948–956
Mattsson JL, Spencer PJ, Albee RR (1996) A performance standard for chemical and functional observation battery examinations of rats. J Am Coll Toxicol 15:239–254
Miyazaki H, Tagawa M (2002) Rate-correction technique for QT interval in long-term telemetry ECG recording in beagle dogs. Exp Anim 51(5):465–75
Miyazaki H, Watanabe H, Kitayama T, Nishida M, Nishi Y, Sekiya K, Suganami H, Yamamoto K (2005) QT PRODACT: sensitivity and specificity of the canine telemetry assay for detecting drug-induced QT interval prolongation. J Pharmacol Sci 99(5):523–9
Omata T, Kasai C, Hashimoto M, Hombo T, Yamamoto K (2005) QT PRODACT: comparison of non-clinical studies for drug-induced delay in ventricular repolarization and their role in safety evaluation in humans. J Pharmacol Sci 99(5):531–41
Pugsley MK, Authier S, Curtis MJ (2008) Principles of safety pharmacology. Br J Pharmacol 154:1382–1399
Redfern WS, Valentin J-P (2011) Trends in safety pharmacology: posters presented at the annual meetings of the Safety Pharmacology Society 2001–2010. J Pharmacol Toxicol Methods 64:102–110
Redfern WS, Carlsson L, Davis AS, Lynch WG, MacKenzie I, Palethorp S, Siegl PKS, Strang I, Sullivan AT, Wallis R, Camm AJ, Hammond TG (2003) Relationships between preclinical cardiac electrophysiology, clinical QT interval prolongation and torsade de pointes for a broad range of drugs: evidence for a provisional safety margin in drug development. Cardiovasc Res 58:32–45
Sanguinetti MC, Mitcheson JS (2005) Predicting drug-hERG channel interactions that cause acquired long QT syndrome. Trends Pharmacol Sci 26(3):119–24
Sanguinetti MC, Tristani-Firouzi M (2006) hERG potassium channels and cardiac arrhythmia. Nature 440:463–469
Sasaki H, Shimizu N, Suganami H, Yamamoto K (2005) QT PRODACT: inter-facility variability in electrocardiographic and hemodynamic parameters in conscious dogs and monkeys. J Pharmacol Sci 99(5):513–22
Schram G, Zhang L, Derakhchan K, Ehrlich JR, Belardinelli L, Nattel S (2004) Ranolazine: ion-channel-blocking actions and in vivo electrophysiological effects. Br J Pharmacol 142:1300–8
Song Y, Shryock JC, Wu L, Belardinelli L (2004) Antagonism by ranolazine of the pro-arrhythmic effects of increasing late INa in guinea pig ventricular myocytes. J Cardiovasc Pharmacol 44:192–9
Strnadova C (2005) The assessment of QT/QTc interval prolongation in clinical trials: a regulatory perspective. Drug Inf J 39:407–433
Tashibu H, Miyazaki H, Aoki K, Akie Y, Yamamoto K (2005) QT PRODACT: in vivo QT assay in anesthetized dog for detecting the potential for QT interval prolongation by human pharmaceuticals. J Pharmacol Sci 99(5):473–86

Toyoshima S, Kanno A, Kitayama T, Sekiya K, Nakai K, Haruna M, Mino T, Miyazaki H, Yano K, Yamamoto K (2005) QT PRODACT: in vivo QT assay in the conscious dog for assessing the potential for QT interval prolongation by human pharmaceuticals. J Pharmacol Sci 99(5):459–71

Trepakova ES, Koerner J, Pettit S, Valentin JP (2009) A HESI consortium approach to assess the human predictive value of non-clinical repolarization assays. J Pharmacol Toxicol Methods 60:45–50

Yamazaki R, Yamada M, Kobayashi K, Kitani S, Shimada C, Shimosato T, Mori T, Suganami H, Yamamoto K (2005) QT PRODACT: inter- and intra-facility variability of the action potential assay using isolated guinea-pig papillary muscles. J Pharmacol Sci 99(5):439–47

Yan GX, Antzelevitch C (1996) Cellular basis for the electrocardiographic J wave. Circulation 93:372–379

Zbinden G (1979) Pharmacological methods in toxicology. Pergamon, Elmsforn, NY

Chapter 12
ICH S8: History and Perspectives

Kenneth L. Hastings

Abstract An important aspect of drug safety evaluation is determination of potential adverse effects on immune function. Drug-induced immune dysfunction can present as increased susceptibility to infections and tumors (especially virally induced), hypersensitivity reactions such as drug allergy and autoimmunity, and various inflammation-like phenomena. Although immunotoxicity test methods have been developed to assess environmental chemicals, these had not been applied systematically in drug development prior to promulgation of guidance documents by EMA and FDA. EMA and FDA guidances/guidelines differed in certain important respects, and ICH S8 was written to resolve these somewhat conflicting approaches. The key issue resolved in ICH S8 was whether functional immunotoxicity assays should be conducted routinely or when there was a cause for concern. An important result of ICH S8 is that drug developers can no longer ignore signs of compound-related adverse effects on immunity. ICH S8 provides a systematic approach to determining the need for immunotoxicity testing and includes discussion on appropriate methodology. Based on current experience with ICH S8, the issue of including immune function parameters in standard toxicity testing remains unresolved and may be addressed in future revisions of the document. In addition, guidance on unintended immunostimulation may be needed based on recent experiences in clinical drug development.

12.1 Introduction

Although immunotoxicology was first identified as a distinct specialty in toxicology by Vos (1977), the study of adverse effects on immune function parallels the emergence of immunology. Richet and Portier first described anaphylaxis in 1902, and Auer, in 1911, made the crucial discovery that this reaction required previous

K.L. Hastings (✉)
Sanofi SA, Bethesda, MD, USA
e-mail: kenneth.hastings@sanofi-aventis.com

J.W. van der Laan and J.J. DeGeorge (eds.), *Global Approach in Safety Testing*, AAPS Advances in the Pharmaceutical Sciences Series 5, DOI 10.1007/978-1-4614-5950-7_12,

exposure to the causative substance (Portier and Richet 1902; Auer 1911). These observations became important in early toxicology with the near concurrent discovery of penicillin and Landsteiner's groundbreaking work with what are now called haptens. Often overlooked when considering the discovery of what was (rightly) considered the miracle drug penicillin was an all-too-common side effect: anaphylaxis (Feinberg et al. 1953). Meanwhile, Landsteiner (1945) demonstrated that small molecular weight chemicals could bind irreversibly to proteins and, when injected into animals, could induce an immune response. Landsteiner's work would be utilized in what should be considered the first immunotoxicology assay—the Draize test for allergic contact dermatitis (ACD) (Draize et al. 1944). In the 1950s, Ovary and colleagues established the link between hapten–protein formation and induction of anaphylaxis. The first method useful in studying (if not predicting) anaphylaxis was developed by Ovary: the passive cutaneous anaphylaxis assay (PCA) (Ovary 1958).

Immune-mediated hypersensitivity reactions are now considered a type of immunotoxicity, but unintended *immunosuppression* has long been considered the more important immunotoxic effect. There were two causes identified early in toxicology: ionizing radiation and certain highly reactive chemicals. Both seemed to share a common mechanism: bone marrow toxicity (Auerbach 1958). That both could be toxic to rapidly dividing cells led to experimental therapies for cancer and to prevent rejection of transplanted organs. Thus, the link between cell proliferation and immunity was understood before many of the discoveries we now take for granted.

By the 1970s, many toxicologists and other biomedical scientists understood that the immune system was like any other: susceptible to insult which would lead to impaired function. But systematic investigation of immunotoxicity began, as is true for many subjects of interest to toxicologists, with a disaster. In 1973, an industrial accident in Michigan led to contamination of milk and milk products with polybrominated biphenyls (PBBs). Subsequent toxicology studies demonstrated PBBs to cause adverse effects on immune function in animals and humans (Bekesi et al. 1978). Other chemicals (such as aflatoxin) were found to have similar adverse effects on immune function (Thaxton et al. 1974). This was the beginning of immunotoxicology as a distinct specialty.

Food contamination is considered to be essentially an environmental problem (with the important exception of intentional adulteration), and it is thus not a surprise that immunotoxicology was developed by scientists working for regulatory agencies such as the US Environmental Protection Agency (US EPA). But it is often forgotten that pharmaceutical scientists were also interested in the subject. In 1978, the US Food and Drug Administration held a meeting on adverse effects of drugs on immunity. One paper from this meeting is of particular interest: Nelson Irey, a pathologist at the Walter Reed Armed Forces Institute of Pathology, for what appears to be the first time, grouped several types of adverse drug reactions into a single category as "immunotoxicity." Irey included penicillin-induced anaphylaxis, α-methyl dopa-induced autoimmunity, radiation and cancer chemotherapy-induced susceptibility to infections and tumors, azathioprine-induced lymphoma-like lymphoproliferative disease in renal transplant patients, and vaccine-related hypersensitivity reactions in this category (Irey 1978). For perhaps the first time, apparently unrelated pathologies were understood to have a common basis: immune system impairment.

Irey's observations did not result in a systematic approach to evaluation of xenobiotics for immunotoxic potential. This would be accomplished by Vos and his colleagues in the Netherlands, and somewhat later by scientists at US EPA, the National Institute of Environmental Health Science (NIEHS), and a few universities and associated research institutes (House and Luebke 2007). The close association between classic immunology and the new science of immunotoxicology is evident in the assays that were developed to detect immunotoxic chemicals. Virtually all of the assays are adaptations of methods that had long been used by immunologists to study basic immunobiology. For example, Cunningham (1965) reported an assay that used sheep red blood cells (SRBC) to study immune responses in rodents. This method, now known as the *plaque assay*, was used by Jerne and colleagues to study mechanisms of immune specificity (Jerne and Nordin 1963; Forni et al. 1980). The plaque assay was adapted by Vos and others to study immune impairment by xenobiotic exposure (Dean et al. 1982; Vos 1977). Just as the sheep red blood cell (SRBC) plaque assay was pivotal to understanding the evolution of immune specificity, so to it would prove to be the most useful general assay for detection of xenobiotic immunotoxicants (Putman et al. 2002; Van der Laan et al. 1997).

The second method of importance is what is generally referred to as the "host-resistance assay." Essentially, rodents are exposed (by various routes depending on the challenge agent) to either infectious agents or tumor cells. In classic immunobiology, these models have been used to study immune responses to infections and cancers. Immunotoxicologists adapted these models to study the effects of xenobiotics on the immune response to these challenge agents (Burleson and Burleson 2008).

The SRBC plaque assay is now generally referred to as the T-dependent antibody response (TDAR) assay and combined with host-resistance assays, constitutes the "gold standard" in immunotoxicology (Luster et al. 1993). There are two important points to consider: both assays are relatively nonspecific (they do not "predict" a specific human health effect) and both are assays for impaired *function*. Both are also cornerstone assays recommended in ICH S8.

12.2 Immunotoxicology and ICH

In the early 1980s, a group of scientists from several institutions developed a series of assays to detect xenobiotic immunotoxicity. The nexus of this effort was the US National Toxicology Program (NTP), and the suite of tests is commonly referred to as the "Tier Assays" (Luster et al. 1988). The overall methodology is familiar to toxicologists: one or more relatively nonspecific, but sensitive, assays are used to screen for effect(s) of concern with follow-on tests to confirm and understand targets of toxicity. In the case of immunotoxicity, a series of studies was conducted to determine "concordance," that is, which assays were most useful in identifying known immunotoxicants (Luster et al. 1992, 1993). Fifty chemicals was evaluated using adaptations of several methods commonly used by immunologists. These included xenobiotic effects on immune system organ weights, cellularity, and histology;

certain clinical pathology parameters (blood cell counts, flow cytometry, serum immunoglobulin concentration); and a set of functional assays (TDAR, delayed-type hypersensitivity, cell-mediated immunity, NK cell activity, and host resistance). The most concordant assay was the TDAR and combined with flow cytometry, detected around 90% of the 50 immunotoxicants (House and Luebke 2007). Two important points should be made: the study lacked *negative* controls, and few drugs were included. But this was an important beginning for immunotoxicology: a reasonable approach to evaluation of potential immunotoxicants could now be recommended.

In 1989, the National Institute of Public Health and the Environment in The Netherlands began what was probably the first systematic evaluation of drug immunotoxicity, including assessment of some not considered likely to be immunomodulatory (e.g., verapamil). An important result of these studies was demonstration that methods developed for immunotoxicity evaluation of environmental chemicals also were useful for drugs (De Waal et al. 1995, 1996, 1997, 1998; Van der Laan et al. 1995, 1996). Studies conducted by the US NTP to evaluate the potential immunotoxicity of drugs for treatment of AIDS produced some evidence of concern (Luster et al. 1991). One, zalcitabine, was found to suppress T-helper cell numbers in cynomolgus monkeys (Taylor et al. 1994). This often overlooked study demonstrated two important points: flow cytometry, conducted in a non-rodent species, could yield clinically relevant results (perhaps explaining why a potent antiretroviral drug could reduce mortality without an increase in the accepted surrogate marker for efficacy), and that an immunotoxicity end point could be obtained in a study that was not designed as a stand-alone immunotoxicology assay.

Immunotoxicology assays, demonstrated by Vos and colleagues to be useful for drugs, would be standardized for evaluation of environmental contaminants and food additives, but not for pharmaceuticals. There are several reasons for this. In drug development, flexibility in study design and end points was considered more appropriate, and protocols developed by the Organization for Economic Cooperation and Development (OECD) were used for general guidance (OECD 1995, 2008). For many drugs, it was difficult to distinguish immunotoxicity from exaggerated pharmacodynamics (e.g., transplant drugs, anti-inflammatory drugs, and cancer chemotherapeutics). In fact, it can be argued that the distinction is arbitrary: unlike environmental chemicals, drugs are evaluated in the context of benefit/risk for a given therapeutic indication (Piccotti et al. 2009). In a study conducted with another HIV drug, didanosine, it was shown that immunotoxic effects could be demonstrated with much longer exposure (6 months) than recommended under EPA guidelines (Phillips et al. 1997). Finally, the most important immunotoxic effect associated with drugs appeared to be *hypersensitivity reactions*, not unintended immunosuppression. The EPA tier system was not designed to detect enhanced immune responses that appeared to be the mechanism of what is commonly referred to as "drug allergy."

Nevertheless, it was understood that immunotoxicity, including unintended immunosuppression, was a potentially important adverse drug reaction. Probably, the first to address this issue was the pharmaceutical industry itself. In a 1988 white

paper, the Pharmaceutical Manufacturers Association (PMA), the forerunner of the US industry organization PhRMA, presented a rationale for incorporation of immunotoxicity end points into drug safety evaluation but cautioned that the EPA tier system was not useful in this context (PMA 1988). Two meetings sponsored by the Drug Information Association (DIA) were important for obtaining consensus on the need for guidance on immunotoxicity evaluation of new drugs (Arlington, Virginia, in 1995 and Montreux, Switzerland, in 1996). Consensus was reached in Montreux on a general approach to immunotoxicity testing with emphasis on use of the TDAR as the best general assay (Van der Laan et al. 1997).

US FDA had included an adaptation of the EPA tier system in the first edition of what is commonly referred to as the "Red Book," but these recommendations applied to new food additives, not drugs (Hinton 2000). The first FDA guidance for immunotoxicity evaluation of drugs came from the Division of Antiviral Drug Products in the Center for Drug Evaluation and Research and was developed to address concerns that AIDS patients should not be given drugs that were immunosuppressant (FDA 1993; Hastings 1996). However, it was clear that the guidance was insufficient and that a more formal document should be written.

Concurrent with FDA guidance development, health authorities in Europe were also concerned that potential immunotoxicity was not being appropriately evaluated as part of drug development. It was the divergence of opinion on a specific point that would lead to ICH S8. The European perspective was that *functional* assays should be conducted to evaluate investigational drugs as part of routine safety evaluation (Putman et al. 2003; Vos and Van Loveren 1998). This position was consistent with the approach taken by EPA: the important parameter was potential adverse effects on immune function and should be the basis for any policy recommendation. This position was clearly justified by the available scientific evidence.

The position taken by FDA was that dedicated immunotoxicity studies might not be necessary if the totality of data from nonclinical (and clinical) studies were properly evaluated. The approach advocated by CDER/FDA was that signs of unintended immunosuppression could be observed and only then would dedicated immunotoxicity studies be useful (Hastings 2002).

There was much discussion in the 1990s on the issue of dedicated versus cause-for-concern studies for immunotoxicity evaluation of investigational drugs. The proximal cause for ICH emerged from these discussions. In 2000, the European Agency for the Evaluation of Medical Products (predecessor to the European Medicines Agency; EMA) published a note for guidance on 28-day rodent toxicity studies, which included an appendix that called for dedicated immunotoxicity studies (EMEA 2000). Draft guidance on immunotoxicity evaluation of drugs by the Japanese Ministry of Health, Labor, and Welfare (JMHLW) advocated a tiered approach consistent with EMA: a functional assay should be conducted as part of standard nonclinical safety assessment. FDA/CDER promulgated a guidance that advocated a cause-for-concern approach (US FDA 2002). Clearly, there was a divergence of opinion—resulting in the need for ICH negotiations.

12.3 Writing a Guidance

ICH S8 is important because it illustrates how scientists with divergent and strongly held opinions can examine existing evidence and produce a document that appropriately addresses a safety issue. At the initial EWG meeting in London (October 2003), all of the parties involved agreed that immunotoxicity was an important issue to be addressed. At the time, CPMP (now CHMP) guidance was dominant: dedicated immunotoxicity studies were needed, whereas for CDER/FDA they *might* be needed depending on available data.

The first task of the immunotoxicology expert working group (EWG) was to determine the approaches in use at the time by the pharmaceutical industry to screen drug candidates for immunotoxic potential. This survey found that although there was considerable variability within industry, most relied on standard nonclinical toxicology studies to detect signs of immunotoxicity (Weaver et al. 2005). Some companies conducted immune function studies such as TDAR but almost always if signs suggestive of immunotoxicity had been observed in nonclinical toxicology studies or if there were other causes for concern. The most important finding from both the survey of industry practices, as well as experience by the regulatory agencies, was that signs of immunotoxicity were often either ignored, considered not relevant to clinical use, or were due to "stress." Thus, the problem did not appear to be lack of immunotoxic signs but failure to appropriately evaluate these. Concern over this particular point was important in formulation of specific guidance in the resulting document.

The second task was to determine the need for dedicated *functional* immunotoxicity studies as part of routine drug development. Although the CPMP NfG *seemed* to require either a TDAR or a combination of flow cytometric analysis of immune cells combined with natural killer (NK) cell activity as part of a 28-day repeat-dose toxicology study in rats, this may have been a false interpretation. In fact, the NfG strongly recommended including immune function end points unless there was a compelling reason no to. In effect, FDA and EMA guidances on immunotoxicology differed in recommended approach, not in whether such determinations were needed. FDA/CDER recommended follow-on immunotoxicity testing if there was a cause for concern, whereas EMA and JMHLW recommended dedicated testing unless there was *no* cause for concern. Thus, there was a basis for consensus: both regulatory authorities agreed on the need for immunotoxicity testing, and both agreed that tests such as the TDAR could be recommended.

Finally, the EWG had to consider the *scope* of the guidance. The FDA Guidance on Immunotoxicology Evaluation of New Drugs included an extensive discussion of phenomena generally referred to as "drug allergy" (US FDA 2002). Many types of drug-associated immunopathies are included in this category, but few test methods could be recommended to determine the potential of investigational drugs to cause these adverse effects. There are many methods to determine the ability of a drug administered by dermal application to cause allergic contact dermatitis (ACD), but these are accepted as adequate by all parties in the EWG. Given the relatively narrow

scope of the issue and the absence of discordance on acceptable methods, the issue of testing for ACD potential was omitted. As for other types of drug allergy, no methods could be recommended. The issue of anaphylaxis was particularly difficult since Japanese regulatory authorities had long required PCA and a related test, active systemic anaphylaxis (ASA), be conducted as part of routine drug evaluation (Udaka 1992). Aside from the issue of whether nonclinical anaphylaxis assays were useful, there were actually very few available data upon which to make recommendations. There are animal models that can be used to determine if adverse reactions consistent with anaphylaxis are in fact immune mediated, but these have not undergone sufficient validation to recommend. Finally, biologic drugs were not considered in discussions. Primarily, this decision was taken because many biologic drugs are either recombinant immune system proteins such as cytokines or are intended to modulate immune function by some other mechanism. The case-by-case approach that forms the basis of ICH S6 was considered adequate.

One issue was considered important: signs of unintended immunostimulation. The reason for including this topic was that such signs could be, and often were, seen in either nonclinical or clinical studies (Pieters 2008; Rock et al. 2010). Although such signs *could* be due to drug-specific antibody or cell-mediated mechanisms, there are other possible causes. The important point was that *any* sign of unintended immunomodulation should be evaluated when observed, whether consistent with immunosuppression or immunostimulation.

Finally, it should be noted that consensus on the issue of dedicated functional immunotoxicity assays was never achieved. The first problem was the dataset key to determining adequacy of current industry practices (Weaver et al. 2005). Although results of standard nonclinical toxicology studies (STS) accurately predicted immunotoxicity for ~90% of evaluated drugs, the actual number (42) was small. Data were inadequate for evaluation of 12 drugs, and 7 were cytotoxic oncolytics judged inappropriate for inclusion in the analysis. Most troubling was the fact that STS did not detect signs of immunotoxicity discovered with six drugs in dedicated immunotoxicity studies (primarily TDAR). Clinical data were not available for evaluating this most important measure of concordance. Thus, although agreement was achieved on the cause-for-concern approach, there was a risk that drug-induced unintended immunosuppression could be undetected.

12.4 ICH S8: The Essentials

The ICH S8 guidance document was negotiated for about 2 years—a remarkably short period of time compared to other safety topics. Immunotoxicity was accepted as an ICH topic in Osaka (November 2003), the first draft was produced in McLean, Virginia, in June 2004, and the pivotal Step 2 document was finalized in Yokohama in November 2004. The Step 4 document was signed in Brussels in May 2005, and the final guidance was published August 23, 2005. FDA promulgated the guidance in April 2006. The objectives of the final document are simple: to recommend

methods to evaluate immunotoxic potential of investigational drugs and to provide a scientifically based algorithm to determine the circumstances in which dedicated nonclinical immunotoxicity testing would be needed. Two linked methods form the structure of the guideline: a cause-for-concern paradigm which informs a weight-of-evidence determination of need for further studies.

The phrase "cause for concern," although not used in the guideline, captures the approach to evaluate need for specific immunotoxicity testing. The following factors should be considered (1) findings in nonclinical toxicology studies, (2) pharmacology of the drug, (3) indication, (4) potential structure–activity relationship(s), (5) pharmacokinetics, and (6) relevant observations in clinical use. This is a holistic approach: *all* relevant data should be evaluated for signs of test article immunomodulation.

The first cause for concern is important because there is no reliance on a specific toxicology study. Consider this in contrast to the requirements promulgated in the US EPA Office of Prevention, Pesticides, and Toxic Substances (OPPTS) *Health Effects Test Guidelines*: *OPPTS 870.7800 Immunotoxicity* (US EPA 1998). The EPA Guideline is very specific: the TDAR should be conducted in a 28-day repeat-dose oral administration rodent study, with the high-dose group given the maximum tolerated dose (MTD). For determination of potential test article effects on NK cells, the length of exposure should be 90 days using the same method of administration. ICH S8, in contrast, recommends that results obtained in *all* nonclinical studies be evaluated. There is a trade-off that should be understood: rather than rely on a single data-dense method (TDAR) in a single species (usually rats), the approach given in ICH S8 relies on signals from multiple nonclinical studies in both rodents and non-rodents. This methodology could be considered "data sparse" in comparison to the EPA approach, but vigilance in study analysis should correct this potential problem. Thus, ICH S8 provides an extensive list of relevant observations that could indicate potential immunotoxicity (including recommendations on histopathology).

Signs of immunotoxicity include alterations in immune tissue weight, cellularity, and histologic appearance, blood immunoglobulin changes, and increased incidence of infections and tumors. Anatomic and biochemical changes can suggest immunosuppression or immunostimulation: either should be evaluated. For example, increased numbers of lymph node and splenic germinal centers may be taken to suggest adverse immunoenhancement, but in combination with increased incidence of infections could, in fact, be indicators of immunosuppression. Tumor findings in rodent carcinogenicity bioassays could suggest immunosuppression if there are no other known relevant mechanisms (such as genotoxicity or hormonal activity).

The usefulness of histopathology was of particular concern. Immunotoxicologists have long debated whether immunosuppression can occur in the absence of histologic changes. S8 emphasizes an approach that toxicologic pathologists have called "enhanced histopathology." In addition, S8 provides a list of tissues that should be specifically evaluated for immune effects. Combined with published "best practices" and the fact that many nonclinical studies would be evaluated, this issue should be considered somewhat resolved (Haley et al. 2005; Kuper et al. 2000; Maronpot 2006). Certainly, there are examples where reliance on histopathology alone, especially when conducted using tissues obtained in a 1-month rodent study, would

fail to detect some immunosuppressant compounds, but this is not the holistic approach given in S8.

The second cause for concern is the pharmacology of the drug. There are at least three categories of drugs that could demonstrate signs consistent with immunotoxicity (1) cancer chemotherapeutics, (2) transplant drugs, and (3) anti-inflammatory drugs. In all three cases, immunotoxicity is likely to be exaggerated pharmacodynamics, and dedicated immunotoxicity studies might not provide useful information. However, this should not be a default assumption. Many standard cancer chemotherapeutic drugs are bone marrow toxins, and medical practice has long taken this into consideration (such as isolating patients to minimize infection risk). However, newer chemotherapeutics with novel molecular targets may not have obvious immunosuppressive activity (e.g., Yang and Moses 2008). In this case, even in the absence of concerning signs in nonclinical toxicology studies, determination of immunotoxic potential could be advisable. Drugs developed to prevent organ transplant rejection are often considered to be "obvious immunotoxins," and end points obtained in dedicated immunotoxicity studies have typically been captured in pharmacology studies. Once again, however, this may not be the case for drugs with unique pharmacodynamic properties. Immunotoxicology studies could be useful to separate wanted pharmacodynamics from unintended immunosuppressive effects.

Anti-inflammatory drugs constitute a special category for consideration. This can be illustrated by an often forgotten episode in the early development of steroid drugs. Clinical trials were conducted in patients with tuberculosis: the potent anti-inflammatory (as well as anabolic) effects of corticosteroids resulted in remarkable resolution of symptoms (Shubin et al. 1959). These effects were, of course, temporary, and patients soon developed serious, often fatal, recurrence of active tuberculosis. Pharmacodynamic activity resulted in fatal immunotoxicity. This episode in drug development was accidentally reproduced when the first anti-inflammatory biologics were used to treat rheumatoid arthritis: patients with occult tuberculosis infection sometimes developed fatal active disease (Dixon et al. 2010). The essential issue in evaluating anti-inflammatory drugs is to determine the therapeutic ratio based on immune system parameters: there is an overlap between immunopharmacology and immunotoxicology. Although therapeutic ratio is an issue with virtually any drug, the issue appears to be especially complex for anti-inflammatory drugs. Steroids can be used in combination with antituberculosis drugs for certain manifestations of the disease where inflammation is an important pathologic feature (Cunha 1995). Antitumor necrosis factor alpha (TNFα) monoclonal antibodies can be used for effective treatment of chronic inflammatory diseases such as rheumatoid arthritis and Crohn's disease if the patient does not have tuberculosis infection (CDC 2004). Adverse immunomodulation has been observed with other anti-inflammatory drugs (especially biologics), and nonclinical methods may be useful on a case-by-case basis (Gourley and Descotes 2008).

Indication is a cause for concern if the drug will be given to patients with impaired immune function (HIV patients, children with congenital immunodeficiency, elderly patients). In this context, immunotoxicity studies are likely to identify hazard, but risk would be determined in clinical trials. There was considerable

debate on the issue of indication and patient population. Although some drugs indicated for treatment of HIV infection had been demonstrated to have immunosuppressive effects in both STS and dedicated immunotoxicity studies, these findings had little, if any, impact on product labels or clinical use. The EWG recognized that infants and children could be especially vulnerable to unintended immunosuppressive effects, but did not make specific recommendations on methods for evaluating this possibility. In fact, other than including possible developmental immunotoxicity as a cause for concern (including in utero exposure), the guidance is otherwise relatively silent on best practices for addressing the issue.

Although structural similarity to known immunotoxic drugs is a cause for concern, this is a complex issue. If a drug exhibits a structural alert when analyzed using an in silico method, it is unlikely that immunotoxicity studies would be needed in the absence of signals seen in in vivo studies. Conversely, if signs of immunotoxicity are observed in toxicology studies and there is a structural alert as well, follow-on immunotoxicity studies should be considered.

The most important pharmacokinetic parameter that could indicate a cause for concern is disposition. If a drug and/or a metabolite accumulates in immune system tissues, this would not be stand-alone cause for concern. However, if there are other findings in toxicology studies consistent with such immune tissue accumulation (such as histopathologic alterations), this would be a cause for concern.

Clinical trial data may indicate cause for concern. There are many clinical findings that could indicate the need for nonclinical immunotoxicity studies. Often these studies would be needed to help establish a link between clinical findings such as increased incidence of pneumonia or urinary tract infections and immunosuppression due to the drug. This is not a rare event and has been seen with both drugs and biologics. For example, proton pump inhibitors appear to increase risk for pneumonia and *Clostridium difficile* infections (Gulmez et al. 2007; Linsky et al. 2010). Anti-adhesion molecule monoclonal antibodies may increase the risk of active JC virus encephalopathy (Bloomgren et al. 2012). More complex are issues such as the potential association between acetaminophen and risk of asthma (Eyers et al. 2011). Some types of immunotoxicity appear to decrease vaccine efficacy (Gelinck et al. 2008; Grandjean et al. 2012).

Finally, the issue of stress was extensively discussed. As noted previously, there is a long and troubling history of drug-associated immune impairment being dismissed as stress-related and not relevant to clinical safety. The complexity of the issue perhaps can be best understood as the conundrum of toxicity-induced stress. If thymic atrophy, for example, is observed in animals demonstrating evidence of toxicity not related to immune function, should this be considered immunotoxicity? There are no simple answers to this question, but the EWG concluded that far too often stress is the default explanation for observed signs of immunotoxicity and that this was not acceptable. Thus, there is the statement in the guidance that if the claim is made that signs of immunotoxicity are due to stress, *compelling* evidence must be provided to support this conclusion. Although the guidance is not specific about what should be considered compelling evidence, the implication is that a simple statement of stress causality would not

be sufficient. The guidance recommends that doses used in STS should be less than MTD in order to minimize potential for stress. The Appendix includes a thorough discussion of stress-related effects which could inform interpretation of STS findings.

The Appendix also includes an extensive discussion of specific immune function assays. Thus, although the guidance does not provide a "recipe" for conduct of studies (i.e., specific requirements), numerous useful points to consider are provided. In this respect, the guidance is somewhat unique. Flexibility in study design, based on various considerations, is recognized as an important factor.

12.5 Maintenance

ICH recognizes that scientific advances influence conduct of studies and that there will be a need to update guidance documents. There are several issues that may necessitate maintenance of ICH S8.

ICH S8 does not address the issue of drug allergy. Although there are few methods that can be recommended, the murine local lymph node assay (LLNA) should be considered appropriate to evaluate the safety of dermal drugs. This assay is validated and generally accepted (Gerberick et al. 2005). Although there are few published data using the LLNA to evaluate the potential of drugs to produce ACD, it is unlikely that inclusion in a revised guidance would be controversial.

Although developmental immunotoxicity is recognized as an important cause for concern, specific guidance is not provided in ICH S8. Since promulgation of ICH S8, there have been important advances in developmental immunotoxicology, and this issue should be addressed as part of maintenance (Holsapple et al. 2005). Although controversy is likely on some key aspects of both study design and need for studies, these issues could be successfully addressed in negotiations. Consideration should also be given to evaluation of the immunotoxic potential of drugs intended for use in the elderly.

Biologic drugs are not in the scope of ICH S8, but consideration should be given to this issue. ICH S6(R1) defers to ICH S8 on some important aspects of drug evaluation: especially important is the issue of infections and tumors associated with biologic immunomodulators. Although recommendations are made in the Appendix of ICH S8 on host-resistance assays, this section could be greatly expanded and could provide useful guidance applicable to biologic drugs.

Advanced techniques such as genomics have been applied in immunotoxicology: it is unclear, however, if guidance is needed. However, this is a rapidly changing area of drug safety evaluation, and consideration should be given to whether certain issues should be addressed. For example, immunomics is a technique that could be useful in assessing biologic drugs for adverse immunogenicity (e.g., autoimmune reactions) (Grainger 2004). It is possible that certain epitopes can be identified for which an induced immune response would be a significant hazard.

Adverse immunostimulation is addressed in ICH S8, but no specific guidance is provided on methods for assessment. As part of maintenance, some methods might be worthy of consideration. For example, in vitro methods such as the minimum acceptable biological effect level (MABEL) assay could be recommended to evaluate the safety of agonist immunomodulators (Horvath and Milton 2009; Stebbings et al. 2007). Genomic techniques could also be useful in this context: identified haplotype risk factors could be used to determine potential of test article to produce adverse effects such as "cytokine storm" and "sterile sepsis" (Luebke et al. 2006).

Finally, the original database used in ICH S8 negotiations should be greatly expanded. Useful data are undoubtedly available, and the question of whether immune function assays should be part of standard drug safety assessment can be reexamined. One of the issues that confounded negotiations on this point was whether immunogen challenge could be incorporated into STS without complicating study interpretation. This issue should be considered resolved: the most important issue that remains is optimum parameters for immunogen challenge (e.g., appropriate dose of KLH). There have been examples of unintended immunosuppression with serious clinical consequences (e.g., proton pump inhibitor association with increased risk of pneumonia and *Clostridium difficile* infection, discussed previously). If these adverse immune effects can be modeled in animals (especially as an addition to STS), a strong recommendation could be made that would be a benefit to public health. In addition to TDAR, potential adverse effects on T-independent antibody response and innate immunity should be considered.

ICH S8 should be considered a success: the drug development process has benefited from guidance provided. It is unusual today for signs consistent with immunotoxicity to be ignored or dismissed as "stress" irrelevant to clinical use. Methods to predict drug allergy are needed, and unintended immunostimulation has emerged as a significant problem. But advances in both areas continue to be reported: immunotoxicology is a vibrant field of research with much promise.

References

Auer J (1911) Lethal cardiac anaphylaxis in the rabbit. J Exp Med 14:476–496

Auerbach C (1958) Radiomimetic substances. Radiat Res 9:33–47

Bekesi JG, Holland JF, Anderson HA, Fischbein AS, Rom W, Wolff MS, Selikoff IJ (1978) Lymphocyte function of Michigan dairy farmers exposed to polybrominated biphenyls. Science 199:1207–1209

Bloomgren G, Richman S, Hotermans C, Subramanyam M, Goelz S, Natarajan A, Lee S, Plavina T, Scanlon JV, Sandrock A, Bozic C (2012) Risk of natalizumab-associated progressive multifocal leukoencephalopathy. N Engl J Med 366:1870–1880

Burleson GR, Burleson FG (2008) Animal models of host resistance. In: Herzyk DJ, Bussiere JL (eds) Immunotoxicology strategies for pharmaceutical safety assessment. Wiley, Hoboken, NJ, pp 163–177

Centers for Disease Control and Prevention (CDC) (2004) Tuberculosis associated with blocking agents against tumor necrosis factor-alpha – California, 2002–2003. Morb Mortal Wkly Rep 53:683–686

Cunha BA (1995) Pulmonary tuberculosis and steroids. Chest 107:1486–1487
Cunningham AJ (1965) A method of increased sensitivity for detecting single antibody-forming cells. Nature 207:1106–1107
De Waal EJ, Timmerman HH, Dortant PM, Kranjc MAM, Van Loveren H (1995) Investigation of a screening battery for immunotoxicity of pharmaceuticals within a 28-day oral toxicity study using azathioprine and cyclosporin A as model compounds. Regul Toxicol Pharmacol 21:327–338
De Waal EJ, De Jong WH, Van der Vleit H, Verlaan B, Van Loveren H (1996) An immunotoxicity screening study on salmeterol in rats. Int J Immunopharmacol 18:523–528
De Waal EJ, Van Loveren H, Vos JG (1997) Practice of tiered testing for immunosuppression in rodents. Drug Info J 31:1317–1323
De Waal EJ, Van der Laan J-W, Van Loveren H (1998) Effects of prolonged exposure to morphine and methadone on in vivo parameters of immune function in rats. Toxicology 129:201–210
Dean JH, Luster MI, Boorman GA, Lauer LD (1982) Procedures available to examine the immunotoxicity of chemicals and drugs. Pharmacol Rev 34:137–148
Dixon WG, Hyrich KL, Watson KD, Lunt M, Galloway J, Ustianowski A, Symmons DPM (2010) Drug-specific risk of tuberculosis in patients with rheumatoid arthritis treated with anti-TNF therapy: results from the British Society for Rheumatology Biologics Register (BSRBR). Ann Rheum Dis 69:522–528
Draize JH, Woodward G, Calvery HO (1944) Methods for the study of irritation and toxicity of substances applied topically to the skin and mucous membranes. J Pharm Exp Ther 61:377–390
European Agency for the Evaluation of Medical Products (2000) Committee for Proprietary Medicinal Products. Note for guidance on repeated dose toxicity. European Agency for the Evaluation of Medical Products, London
Eyers S, Weatherall M, Jefferies S, Beasley R (2011) Paracetamol in pregnancy and the risk of wheezing in offspring: a systematic review and meta-analysis. Clin Exp Allerg 41:482–489
Feinberg SM, Feinberg AR, Moran CF (1953) Penicillin anaphylaxis, nonfatal and fatal reactions. J Am Med Assoc 152:114–119
Forni L, Coutinho A, Köhler G, Jerne NK (1980) IgM antibodies induce the production of antibodies of the same specificity. Proc Natl Acad Sci USA 77:1125–1128
Gelinck LBS, Van der Bijl AE, Visser LG, Huizinga TWJ, Van Hogezand RA, Rijkers GT, Kroon FP (2008) Synergistic immunosuppressive effect of anti-TNF combined with methotrexate on antibody responses to the 23 valent pneumococcal polysaccharide vaccine. Vaccine 26:3528–3533
Gerberick FG, Ryan CA, Kern PS, Schlatter H, Dearman RJ, Kimber I, Patlewicz GY, Basketter DA (2005) Compilation of historical local lymph node data for evaluation of skin sensitization alternative methods. Dermatitis 16:157–202
Gourley I, Descotes J (2008) Bridging immunotoxicology to clinical drug development. In: Herzyk DJ, Bussiere JL (eds) Immunotoxicology strategies for pharmaceutical safety assessment. Wiley, Hoboken, NJ, pp 375–384
Grainger DJ (2004) Immunomics: principles and practice. IRTL Rev 2:1–6
Grandjean P, Andersen EW, Budtz-Jørgensen E, Nielsen F, Mølbak K, Weihe P, Heilmann C (2012) Serum vaccine antibody concentrations in children exposed to perfluorinated compounds. J Am Med Assoc 307:391–397
Gulmez SE, Holm A, Frederiksen H, Jensen TG, Pedersen C, Hallas J (2007) Use of proton pump inhibitors and the risk of community-acquired pneumonia. Arch Intern Med 167:950–955
Haley P, Perry R, Ennulat D, Frame S, Johnson C, Lapointe J-M, Nyska A, Snyder PW, Walker D, Walter G (2005) STP position paper: best practices guideline for the routine pathology evaluation of the immune system. Toxicol Pathol 33:404–407
Hastings KL (1996) Regulatory issues of importance to developing anti-HIV therapeutics. Toxicol Pathol 24:278–280
Hastings KL (2002) Implications of the new FDA/CDER immunotoxicology guidance for drugs. Int Immunopharmacol 2:1613–1618

Hinton DM (2000) US FDA "Redbook II" immunotoxicity testing guidelines and research in immunotoxicity evaluations of food chemicals and new food proteins. Toxicol Pathol 28:467–478

Holsapple MP, Burns-Naas LA, Hastings KL, Ladics GS, Lavin AL, Makris SL, Yang Y, Luster MI (2005) A proposed testing framework for developmental immunotoxicology. Toxicol Sci 83:18–24

Horvath CJ, Milton MN (2009) The TeGenero incident and the Duff report conclusions: a series of unfortunate events or an avoidable event? Toxicol Pathol 37:372–383

House RV, Luebke RW (2007) Immunotoxicology: thirty years and counting. In: Luebke R, House R, Kimber I (eds) Immunotoxicology and immunopharmacology. CRC, Boca Raton, FL, pp 3–20

Irey NS (1978) Drug adverse reaction reports related to immunotoxicity. In: Asher IM (ed) Inadvertent modification of the immune response: the effects of foods, drugs, and environmental contaminants. US Food and Drug Administration, Washington, DC, pp 140–143

Jerne NK, Nordin AA (1963) Plaque formation in agar by single antibody-producing cells. Science 140:405

Kuper CF, Harleman JH, Richter-Reichelm HB, Vos JG (2000) Histopathologic approaches to detect changes indicative of immunotoxicity. Toxicol Pathol 28:454–466

Landsteiner K (1945) The specificity of serological reactions. Harvard University Press, Cambridge, MA

Linsky A, Gupta K, Lawler EV, Fonda JR, Hermos JA (2010) Proton pump inhibitors and risk for recurrent *Clostridium difficile* infection. Arch Intern Med 170:772–778

Luebke RW, Holsapple MP, Ladics GS, Luster MI, Selgrade MJ, Smialowicz RJ, Woolhiser MR, Germolec DR (2006) Immunotoxicogenomics: the potential of genomics technology in the immunotoxicity risk assessment process. Toxicol Sci 94:22–27

Luster MI, Munson AE, Thomas PT, Holsapple MP, Fenters JD, White KL Jr, Lauer LD, Germolec DR, Rosenthal GJ, Dean JH (1988) Development of a testing battery to assess chemical-induced immunotoxicity: National Toxicology Program's guidelines for immunotoxicity evaluation in mice. Fundam Appl Toxicol 10:2–19

Luster MI, Rosenthal GJ, Cao W, Thompson MB, Munson AE, Prejean JD, Shopp G, Fuchs BA, Germolec DR, Tomaszewski JE (1991) Experimental studies of the hematologic and immune system toxicity of nucleoside derivatives used against HIV infection. Int J Immunopharmacol 13(S1):99–107

Luster MI, Portier C, Pait DG, White KL Jr, Gennings C, Munson AE, Rosenthal GJ (1992) Risk assessment in immunotoxicology I: sensitivity and predictability of immune tests. Fundam Appl Toxicol 18:200–210

Luster MI, Portier C, Pait DG, Rosenthal GJ, Germolec DR, Corsini E, Blaylock BL, Pollock P, Kouchi Y, Craig W, White DL, Munson AE, Comment CE (1993) Risk assessment in immunotoxicology II: relationships between immune and host resistance tests. Fundam Appl Toxicol 21:71–82

Maronpot RR (2006) A monograph of histomorphologic evaluation of lymphoid organs. Toxicol Pathol 34:409–696

Organisation of Economic Cooperation and Development (1995) Guideline for testing chemicals. Method No. 407: Repeated dose oral toxicity – rodents: 28-day or 14-day study. OECD, Paris (Updated 2008)

Ovary Z (1958) Immediate reactions in the skin of experimental animals provoked by antibody-antigen interaction. Prog Allerg 5:459–508

Pharmaceutical Manufacturers Association (1988) Immunotoxicology Task Force, Report to the Drug Safety Subsection Steering Committee on Immunotoxicology and the Pharmaceutical Industry. PMA, Washington, DC

Phillips KE, McCay JA, Brown RD, Musgrove DL, Meade BJ, Butterworth LF, Wilson S, White KL Jr, Munson AE (1997) Immunotoxicity of 2′,3′-dideoxyinosine in female B6C3F1 mice. Drug Chem Toxicol 20:189–228

Piccotti JR, Lebrec HN, Evans E, Herzyk DJ, Hastings KL, Burns-Naas LA, Gourley IS, Wierda D, Kawabata TT (2009) Summary of a workshop on nonclinical and clinical immunotoxicity assessment of immunomodulatory drugs. J Immunotoxicol 6:1–10

Pieters R (2008) Systemic hypersensitivity. In: Herzyk DJ, Bussiere JL (eds) Immunotoxicology strategies for pharmaceutical safety assessment. Wiley, Hoboken, NJ, pp 241–256

Portier P, Richet C (1902) De l'action anaphylactique de certain venins. C R Soc Biol 54:170–172

Putman E, Van Loveren H, Bode G, Dean J, Hastings K, Nakamura K, Verdier F, Van der Laan J-W (2002) Assessment of immunotoxic potential of human pharmaceuticals, a workshop report. Drug Info J 36:417–427

Putman E, Van der Laan J-W, Van Loveren H (2003) Assessing immunotoxicity: guidelines. Fundam Clin Pharmacol 17:615–626

Rock KL, Latz E, Ontiveros F, Kono H (2010) The sterile inflammatory response. Annu Rev Immunol 28:321–342

Shubin H, Lambert RE, Heiken CA, Sokmensuer A, Glaskin A (1959) Steroid therapy and tuberculosis. J Am Med Assoc 170:1885–1890

Stebbings R, Findlay L, Edwards C, Eastwood D, Bird C, North D, Mistry Y, Dilger P, Liefooghe E, Cludts I, Fox B, Tarrant G, Robinson J, Meager T, Dolman C, Thorpe SJ, Bristow A, Wadhwa M, Thorpe R, Poole S (2007) "Cytokine storm" in the phase I trial of monoclonal antibody TGN1412: better understanding the causes to improve preclinical testing of immunotherapeutics. J Immunol 179:3325–3331

Taylor LD, Binienda Z, Schmued L, Slikker W Jr (1994) The effect of dideoxycytidine on lymphocyte subpopulations in nonhuman primates. Fundam Appl Toxicol 23:434–438

Thaxton JP, Tung HT, Hamilton PB (1974) Immunosuppression in chickens by aflatoxin. Poult Sci 53:721–725

Udaka K (1992) Cellular and humoral mechanisms of immunotoxicological tissue manifestations induced by immunotoxic drugs. Toxicol Lett 64/65:93–100

US Environmental Protection Agency (1998) Health effects test guidelines: OPPTS 870.7800 Immunotoxicity (revised 2006). US EPA, Washington, DC

US Food and Drug Administration (1993) Points to consider in the immunotoxicology evaluation of new drugs intended for the treatment of HIV infection and related disorders. Division of Antiviral Drug Products, Center for Drug Evaluation and Research, US FDA, Washington, DC

US Food and Drug Administration (2002) Guidance for industry: immunotoxicology evaluation of investigational new drugs. US FDA, Washington, DC

Van der Laan J-W, Krajnc EI, Krajnc-Franken MAM, Van Loveren H (1995) Immunotoxicological screening of morphine and methadone in an extended 28 day study in rats. Int J Immunopharmacol 17:535–543

Van der Laan J-W, Timmerman H, Van Loveren H (1996) Comparison of the in vivo effects of morphine and methadone on natural killer cell activity in spleen, peritoneal cavity, and lungs in rats. Int J Immunopharmacol 18:401–407

Van der Laan J-W, Van Loveren H, Vos JG, Dean JH, Hastings K (1997) Immunotoxicity of pharmaceuticals: current knowledge, testing strategies, risk evaluation & consequences for human health. Drug Info J 31:1301–1306

Vos JG (1977) Immune suppression as related to toxicology. CRC Crit R Toxicol 5:67–101

Vos JG, Van Loveren H (1998) Experimental studies on immunosuppression: how do they predict for man? Toxicology 129:13–26

Weaver JL, Tsutsui N, Hisada S, Vidal J-M, Spanhaak S, Sawada J, Hastings KL, Van der Laan J-W, Van Loveren H, Kawabata T, Sims J, Durham SK, Fueki O, Matula TI, Kusunoki H, Ulrich P, Nakamura K (2005) Development of the ICH guidelines for immunotoxicology evaluation of pharmaceuticals using a survey of industry practices. J Immunotoxicol 2:171–180

Yang L, Moses HL (2008) Transforming growth factor β: tumor suppressor or promoter? Are host immune cells the answer? Cancer Res 68:9107–9111

Chapter 13
ICH S9: Nonclinical Evaluation of Anticancer Pharmaceuticals: A Perspective from Regulators on the Development of the Guideline*

John K. Leighton, Klaus Olejniczak, and Hiroshi Onodera

Abstract The development of a guideline for nonclinical testing strategies for anticancer drugs and biologicals was initiated by the International Conference on Harmonisation (ICH) in 2007. The rationale for developing this guideline was that separate regional guidelines were being or had been developed. By nature, ICH guidelines tend to describe regulatory recommendations rather than the underlying rationale of the recommendations. The purpose of this chapter is not to discuss the document per se but to describe the perspective of regulators on some of the topics discussed during the deliberations in developing the ICH S9 guideline, focusing on major changes to drug development compared to past practices, and to illustrate the principles underlying the recommendations and alternative views that were considered.

13.1 Background

Over the past decades, approaches to the nonclinical development of anticancer pharmaceuticals have been independently discussed and developed in Europe, the USA, and Japan. The nonclinical approaches were not agreed on across product

*This article reflects the personal opinions of the authors and does not necessarily reflect the organizations they represent.

J.K. Leighton (✉)
Office of Oncology Drug Products, Center for Drug Evaluation and Research,
US Food and Drug Administration, Silver Spring, MD, USA
e-mail: john.leighton@fda.hhs.gov

K. Olejniczak
Scientific Director in the Federal Institute for Drugs and Medical Devices (BfArM),
Head Geno - and Reproductive Toxicity Unit (retired), Berlin, Germany

H. Onodera
Pharmaceuticals & Medical Devices Agency, Tokyo, Japan

J.W. van der Laan and J.J. DeGeorge (eds.), *Global Approach in Safety Testing*, AAPS Advances in the Pharmaceutical Sciences Series 5, DOI 10.1007/978-1-4614-5950-7_13,

classes such as biologics and drugs. The available disharmonized guideline resulted in inefficient use of animal resources and ineffective drug development in a critical area of human health. In the USA, nonclinical recommendations for cytotoxic drugs were originally developed in collaboration with the US National Cancer Institute in the 1970s and early 1980s (Prieur et al. 1973; Lowe and Davis 1987). At this time, it appeared that there was little commercial interest in developing drugs for this therapeutic area. By the early 1990s, with growing interest in this field, there was recognition that these recommendations needed to be updated, and the FDA's Center for Drug Evaluation and Research (CDER) clarified their regulatory perspective on anticancer drug development (DeGeorge et al. 1998). Subsequently, development of a guideline for nonclinical oncology drug (small molecule) development was initiated in 2001.

The scope of the FDA's developing guideline changed with the merger of some regulatory functions of FDA's Center for Biological Evaluation and Research into CDER and the recognition of different approaches to safety testing for small molecules and biotechnology-derived pharmaceuticals. For example, as discussed by the Oncology Drug Advisory Committee (ODAC 2006), for small molecules, a toxicology study of 1 month duration in rodents and nonrodents was generally sufficient to initiate a phase I clinical trial and allow for continued clinical dosing as long as patients were benefiting and toxicities were considered acceptable. However, for biotechnology-derived pharmaceuticals with long half-lives, a toxicology study of up to 3 months duration or a study based on the proposed duration of clinical dosing (≥1:1 dosing) in nonhuman primates was sufficient to initiate clinical dosing. Longer term toxicology studies may have been needed to be ongoing to continue clinical dosing, and patients could continue beyond the duration of toxicological support on a case-by-case basis. This example highlighted the need to harmonize the recommendations for small molecules and biotechnology-derived pharmaceuticals or to understand the scientific basis for the different recommendations.

In the European Union (EU), the Safety Working Party (SWP) of the Committee for Medicinal Products for Human Use (CHMP) had developed a guideline for anticancer drug development for the European Union (EMA 1998). The guideline was primarily devoted to cytotoxic/cytostatic drugs that are presumed to have a direct effect on tumor cells. While it focused on the development of single drug treatment, studies to support the clinical development of combinations of anticancer drugs, nonclinical testing to investigate pharmacodynamic, kinetic, and toxicological interactions was also encouraged. The guideline aimed at formulating recommendations for pharmacodynamic investigations and the requirements for toxicological studies prior to phases I, II, and III clinical trials as well as marketing applications. This guideline was withdrawn with the adoption of ICH S9.

The Ministry of Health, Labour and Welfare (MHLW) of Japan was developing nonclinical guideline to address various mechanisms of anticancer therapy but did not include biologics in its scope. Thus, there was substantial concern that, when those guidelines would have been completed, there would not have been a harmonized approach in Japan for nonclinical development of drugs for the treatment of patients with cancer.

The development of the International Conference on Harmonisation (ICH) guideline ICH S9: Nonclinical Evaluation for Anticancer Pharmaceuticals was initiated by the ICH, bringing together representatives from the pharmaceutical industries and regulatory authorities from Japan, the EU, and the USA. (Note that the term "pharmaceutical" is used throughout the guideline and in this chapter to denote both drugs and biotechnology-derived products; where necessary, a distinction is made.) The purpose of the ICH is to discuss scientific and technical aspects for development and registration of pharmaceuticals in order to reduce duplicate testing in the research and development of new pharmaceuticals. The first meeting of the ICH S9 Expert Working Group (EWG) was held in October 2007. In addition to ICH members, observers and interested parties to the process included Health Canada, Swissmedic, and the Biotechnology Industry Organization. As described in the concept paper approved by the ICH Steering Committee, the rationale for developing ICH guideline was the existence of an EU guideline on cytotoxic drugs and separate development of broader guideline for anticancer drugs and biologicals in the USA and Japan (Final Concept Paper 2007). After several meetings, a Step 2 draft guideline prepared by the ICH S9 Expert Working Group, and approved by the Steering Committee, was published by regional regulatory authorities. After considering the public comments received, a final document was signed by regulatory authorities and approved for regional publication by the ICH Steering Committee in October 2009. The guideline is available from the ICH Web site and from regional authorities.

Several points need to be made in regard to the actual writing of the guideline. First, each region brought a well-developed perspective to the discussion, which assisted in the discussions, thus reducing the timeline from the first meeting to the final (Step 4) document. Second, a fairly comprehensive document was available to the EWG as an initial working document; initial meetings were devoted to trimming the document to meet the goal of the EWG to produce a guideline outlining recommendations, not an educational guideline. One party in specific (FDA) had noted that some sponsors were not adept in preparing for an Investigational New Drug Application (IND) filing and initially provided a document that included information to assist those sponsors in preparing an IND, such as details in toxicology study design. The EWG removed much of the "educational" aspects of the guideline as not consistent with the approach of other ICH guidelines. In reading ICH S9, it is important to understand that the EWG avoided certain terms as much as possible, such as "need, needed, shall, must, recommend, required," as not appropriate for an ICH guideline. The EWG also sought to avoid the phrase "if feasible" as studies are sometimes feasible but not scientifically justified. Thus, the EWG used other terms such as "warranted" or "not warranted" to reflect the concepts to be communicated by the EWG, and these terms appear frequently in the guideline. Finally, the ICH S9 EWG met several times with the ICH M3(R2) and ICH S6 EWGs that were meeting concurrently with the ICH S9 EWG in order to avoid inconsistencies among documents. To reduce future maintenance of ICH S9, references are made to those documents as appropriate.

13.2 Major Accomplishments and Discussion Topics of the ICH S9 Expert Working Group

13.2.1 Scope of the Guideline

The scope of the guideline as published at Step 2 was similar in concept in many ways to that of the Step 4 document, both in terms of what clinical development programs are included and what is excluded, that is, what products would fall under the scope of ICH S9 or ICH M3(R2). For example, the S9 guideline covers both small molecule drugs and biotechnology-derived pharmaceuticals but excludes radiopharmaceuticals and vaccines. The rationale for the exclusion is that the nonclinical development programs for these types of molecules would be different than those of "traditional" pharmaceuticals and thus unnecessarily complicate an already difficult task of harmonization.

After the completion of the Step 2 document in November 2008, the EWG spent considerable time discussing the wording of the scope based on public comments received at this Step. At this stage, the guidance was to be applied to pharmaceuticals intended to treat cancer in patients with late stage or advanced disease but was not intended for pharmaceuticals used to treat patients with long life expectancy. Public comments requested clarity around the intended population and requested a definition of "long life expectancy," for example, specifying a potential life expectancy of 3 years. A second comment suggested defining long life expectancy as 80% survival at 5 years. Other comments requested a specific reference to the stage of disease, for example, stage III and stage IV metastatic disease, or that the word "incurable" be added. It is perhaps understandable that clarity around these terms was requested in that potential regional differences in the interpretation of the scope could significantly affect drug development timelines.

In response to the public comments, the EWG attempted to provide clarity to the intended patient population identified in the scope. For example, the EWG considered revisions such as pharmaceuticals being studied for "serious and life-threatening malignancies, which have failed available therapy, or for whom no other therapy exists" and "in patients with metastatic or locally advanced disease and serious and life-threatening hematologic malignancies." The latter language was initially the preferred language of the EWG post-Step 2 and was proposed to the Steering Committee at Step 4 in Yokohama in June 2009. However, the Steering Committee rejected this proposal. Some members of the Steering Committee referred to the concept paper and business plan and the possibility that the proposed language could be an expansion of the guidance beyond originally agreed to in those documents. Other members of the Steering Committee expressed concern that some patients with early-stage disease may receive prolonged treatment without adequate toxicological support (e.g., early-stage breast cancer). Some members of the EWG considered this unlikely as clinical trials for anticancer

pharmaceuticals are almost always initially done in patients with advanced disease that has failed available therapy and that clinical safety data could be used with available toxicology data to support trials in patient populations with less advanced disease without the need for additional toxicology studies. The language originally proposed to the Steering Committee is also used in labeling to describe the patient population of some approved anticancer products.

To address the concerns of the Steering Committee, the EWG discussed alternatives to try to reach consensus on the scope. For example, going back to the original Step 2, language was considered, but this could possibly lead to regional disharmony in applying ICH S9. The EWG also discussed proposals to define or limit the scope to pharmaceuticals intended for patients with disease that has failed available therapy, or where no therapy existed, or where clinical development is initially performed in patients whose disease is refractory to available therapy, or have life-threatening disease where no therapy exists. The proposal included a recommendation that when moving investigations beyond this initial patient population, for example, when a drug is studied in patients with curative intent, long expected survival, or as adjuvant therapy, then the need for additional nonclinical studies would depend upon the available nonclinical and clinical data and the nature of the toxicities observed.

Some EWG members thought these proposals lacked flexibility and interpreted these proposals as suggesting that moving beyond the patient population typically studied in phase I, to phase II or phase III, would lead to the need for additional nonclinical toxicology studies as described in ICH M3(R2). The question of what constituted available therapies was raised: If four or five similar therapies existed for a particular disease, would all therapies need to be tried before an investigational pharmaceutical would fall under the scope of ICH S9? In addition, early-stage planning of a nonclinical program of development could be difficult if some of these concepts (e.g., long expected survival) were to be incorporated into the scope.

In the end, consensus was reached at Step 4, using language similar to that of Step 2, replacing "pharmaceuticals that are only intended to treat cancer in patients with *late stage or advanced disease*" with "*serious and life-threatening malignancies.*" Further, the scope outlined in Step 4 refines "long life expectancy." A key question is whether the principles of ICH S9 or ICH M3(R2) would apply to a particular development program. The scope of ICH S9 addresses this question by stating that the recommendations for and timing of additional nonclinical studies depend upon the available nonclinical and clinical data and the nature of the toxicities observed and did not include reference to curative intent. This statement implies that most development programs for anticancer drugs will initially take place in the setting where therapeutic options may be limited. While not providing specific recommendations, the EWG recognized that moving beyond this initial setting may be possible without additional nonclinical studies on a "case-by-case" basis and that information from the clinical program should inform on this decision.

13.2.2 Role of Pharmacology Investigations in Anticancer Pharmaceutical Development

The EWG discussed in some depth the level of detail and type of pharmacology investigations that were needed to support early development and marketing applications. The EWG discussed whether the assessment should include investigations in specific tumor-derived cell lines in vitro and in xenograft models. For example, if a drug is intended for the treatment of patients with lung cancer, would studies in cell lines derived from lung tumors be needed to support an initial investigation or marketing application? In the end, the consensus of the EWG was that with currently available cell lines, there is not a direct one-to-one concordance between the tumor origin of an in vitro cell line and clinical outcome, so such studies would not be specifically needed; hence, the guideline states that "the pharmaceutical need not be studied using the same tumor types intended for clinical evaluation."

The EWG discussed level of detail or investigations in understanding the mechanism of action of a pharmaceutical was the company's responsibility, but that some rationale should be put forward to justify the clinical trial. It was recognized that a complete understanding of a pharmaceutical was unlikely at early stages of development or even at the time of submission of the marketing application; thus, the level and timing of investigations were left mostly to the discretion of the sponsoring company. However, for biotechnology-derived products, the importance of pharmacology studies in selecting a relevant model, as discussed in ICH S6 (since replaced by ICH S6(R1) 2011), should also be considered.

13.2.3 Duration of Nonclinical Studies to Support Clinical Development

In a significant departure from past practice, the duration and timing of chronic toxicology studies for anticancer pharmaceuticals has evolved. The practice of submitting long-term toxicology study of 6 months' duration in rodents and nonrodents with the marketing application was changed to 3 months to be submitted prior to phase 3. FDA had collected data for about a 6–7-year period to understand how findings from 6-month studies were used; for example, did findings inform clinical monitoring and affect approval recommendations or subsequent clinical investigations in other patient populations? The FDA reported to the EWG that it had no examples to support the need for 6-month studies. Prior to accepting this recommendation, all parties in the EWG consulted with their members and discussed the utility of the current approach of requiring 6-month studies with a marketing application, looking for specific examples where such studies affected clinical development or recommendations. From the response of the EWG, few examples were provided, and it was obvious that long-term toxicology studies submitted with the marketing application had little utility in the course of clinical development and thus the proposal was accepted.

Table 13.1 Examples of treatment schedules for anticancer pharmaceuticals (drugs and biologicals) to support initial clinical trials

Clinical schedule	Examples of nonclinical treatment schedule[a–d]
Once every 3–4 weeks	Single dose
Daily for 5 days every 3 weeks	Daily for 5 days
Daily for 5–7 days, alternating weeks	Daily for 5–7 days, alternating weeks (2-dose cycles)
Once a week for 3 weeks, 1 week off	Once a week for 3 weeks
Two or three times a week	Two or three times a week for 4 weeks
Daily	Daily for 4 weeks
Weekly	Once a week for 4–5 doses

[a]Table 13.1 describes the dosing phase. The timing of the toxicity assessment(s) in the nonclinical studies should be scientifically justified based on the anticipated toxicity profile and the clinical schedule. For example, both a sacrifice shortly after the dosing phase to examine early toxicity and a later sacrifice to examine late onset of toxicity should be considered

[b]For further discussion regarding flexibility in the relationship of the clinical schedule and the nonclinical toxicity studies, see Sect. 3.3 (of the S9 guideline)

[c]The treatment schedules described in the table do not specify recovery periods (see Sect. 2.4 of the ICH S9 guideline and Note 1 regarding recovery)

[d]The treatment schedules described in this table should be modified as appropriate for molecules with extended pharmacodynamic effects, long half-lives, or potential for anaphylactic reactions. In addition, the potential effects of immunogenicity should be considered (ICH guidelines: S6)

The EWG discussed possible scientific rationales that would indicate that a difference in duration of toxicity testing for small molecules and biotechnology-derived pharmaceuticals might be warranted either to initiate a clinical trial or to support marketing. The EWG concluded that the same principles be applied to small molecules and to biotechnology-derived pharmaceuticals. For example, it was noted that some small molecules have a long half-life (e.g., liposome-encapsulated drugs, drugs that bind tightly to serum proteins). Thus, the types of studies needed to support pharmaceutical development programs should be based on sound scientific judgment, taking into account the general recommendations as outlined in Table 1 of the guideline (see Table 13.1).

13.2.4 Dosing Levels in Nonclinical Safety Studies

In general, anticancer drugs developed to date do not have a safety margin, and usually some toxicity in clinical use is anticipated and needs to be managed. For this reason, and since the start dose is based on toxicity, the EWG concluded that defining a No Observed Adverse Effect Level/No Observed Effect Level (NOAEL/NOEL) was not considered essential. Resources should not be dedicated, and toxicology studies repeated, simply to define the NOAEL/NOEL. The rationale behind this concept is another distinguishing feature of anticancer pharmaceutical development.

13.2.5 Defining a "Cytotoxic" Compound by Function

It is recognized that most anticancer therapeutics are "cytotoxic" to one degree or another, and as such, in this context the term is not specific. For this reason, the ICH S9 guideline avoids the use of the term "cytotoxic." The guideline instead refers to a functional capacity of a pharmaceutical to target rapidly dividing cells (e.g., crypt cells, bone marrow) and that are genotoxic; pharmaceuticals in this class are exempt from the need for embryofetal developmental (EFD) toxicity studies as these compounds are either teratogenic or lethal to the developing fetus. The EWG did not address other situations, for example, when pharmaceuticals targeted rapidly dividing cells but were *not* genotoxic, as there was no database to support a conclusion that pharmaceutics in this class are, or are not, teratogenic.

13.2.6 Basis for Reproductive Toxicology Testing

While available information is limited, there is some indication that for some cancers, first detection is at the time of pregnancy diagnosis. For this reason, the ICH S9 guideline focuses on the need for embryofetal development studies of the core battery described by ICH S5(R2) (2002) guideline. The rationale for this approach was to understand the risk to the fetus from unintended exposure if a diagnosis of cancer occurs during early pregnancy. While the entire battery provides important information, for patients with cancer, the EWG consensus was that providing the EFD study alone was sufficient for this patient population. Fertility and pre- and postnatal development studies are not recommended. If pharmaceuticals are to be used in other patient populations, or in the adjuvant setting, then other guidelines would become relevant.

The rationale for not requesting a second embryofetal toxicity study is that if the first is positive, there is no need to confirm a positive finding. In some non-oncology therapeutic areas, there may be a need for a second study to get some idea of a toxic dose and therapeutic dose. Since anticancer drugs are dosed to toxicity in nonclinical studies, and to a maximum tolerated dose in clinical studies, a safety margin is unlikely. Thus, a study in a second species is "not warranted."

13.2.7 Clarifying the Need for Stand-Alone Safety Pharmacology Studies

Safety pharmacology studies investigate functional effects on vital organ function, primarily cardiovascular, central nervous system, and respiratory. Of particular importance is the effect on cardiovascular due to the potential for life-threatening consequence from impairment of this system. The EWG discussed the importance of these studies and concluded that stand-alone studies are not essential to initiate clinical studies as sufficient patient protection is in place with clinical monitoring of cardiovascular function (see ICH E14).

13.2.8 Setting the Start Dose for First Administration in Humans

The EWG discussed several approaches to setting the first in human start dose and concluded that many approaches could be acceptable. The EWG agreed that while it is not likely that most patients will receive a therapeutic benefit while in a phase 1 trial, subtherapeutic dosing should be minimized. In the past, the standard approach for setting a start dose for small molecule drugs was using 1/10th the STD_{10} or that dose that is severely toxic to 10% of rodents (DeGeorge et al. 1998; EORTC 1985). In this case, severely toxic does not necessarily equate to lethality. The EWG agreed that this approach was and could still be useful. Other approaches considered by the EWG were less formulaic, using all the available data, an approach that is common with biotechnology-derived products. This approach was considered more challenging to adopt for small molecules, perhaps leading to greater uncertainty in preparing an initial clinical plan, and it remains to be determined how this will work in practice. Thus, it was thought best by the EWG to provide as much flexibility to sponsors while maintaining patient safety, the approach reflected in the guidance.

13.2.9 The Need for Recovery Groups in All Toxicology Studies

The Step 2 document included language in the General Toxicology Section regarding the need for a recovery (non-dosing) period at the end of the study. In this draft document, the expectation for inclusion of recovery groups to support the initial phase 1 clinical trial was rather definitive. There was also an expectation that progression of toxicity be evaluated. A complete reversibility of findings was not expected; for example, testicular toxicity may not recover within the usual time frame of a 2-week recovery period often used for 1-month toxicology studies.

The EWG had extensive discussion on this topic in responding to public comments while preparing a Step 4 guideline. The EWG noted that there were few, if any, examples of a novel toxicity appearing after the dosing phase. The EWG also decided to provide more flexibility on the inclusion of recovery animals, providing examples where these groups may not be necessary. It was noted that toxicological pathologists were not in complete agreement on the ability to identify lesions that may not recover and that the public literature on this topic was sparse, making the ability to scientifically justify noninclusion of recovery groups difficult. It is clear that findings such as necrosis are not reversible, even if this is not reported upon histopathological examination after the recovery period. It was also reported that some parties do use this data in clinical trial design to determine whether dose interruption or dose decrease may be more appropriate if a particular toxicity is observed. The lack of consensus on this topic was considered a serious deficiency but was included to give sponsors the ability to make a justification. It was also recognized that at some point in the future, a consensus may be reached as to which lesions are reversible and which may require further study.

13.2.10 Integrating Clinical and Nonclinical Data into a Safety Assessment to Support Changes in the Clinical Schedule

Industry representatives to the EWG, and some regulatory parties, thought this topic would be very valuable in clarifying the need, or lack thereof, for additional nonclinical studies to support changes in the clinical dosing schedule, including the clinical dosing schedule proposed during drug development before the first patient had been treated. The rationale for including Sects. 3.3 (initial clinical trials) and 3.4 (continued clinical development, i.e., where some clinical data exist) in the ICH S9 Guideline to address this topic was the lack of clarity and uncertainty in regulatory acceptance of a change in clinical schedule without supporting nonclinical data. This lack of clarity could possibly lead to unnecessary studies and likely increased animal use for little additional information. All parties agreed that the ideal nonclinical program would use a schedule in nonclinical studies similar to that proposed clinically. However, as many of the industry representatives to the EWG pointed out, the complicated nature of pharmaceutical development does not often lend itself to the ideal and that development programs may often change. For example, a drug may be considered for intravenous administration, but new formulations may make oral dosing feasible. After discussion by the EWG, specific factors are provided for consideration in the guideline (Sect. 3.3) to assist in whether additional nonclinical studies would be useful.

13.2.11 Addressing Photosafety Testing

The topic of photosafety testing was incorporated after the Step 2 document was published, in response to public comments received about the Step 2 document in order to address this emerging topic. The ICH S9 EWG discussed various approaches to photosafety testing from “no studies were needed” to “follow the recommendations outlined in ICH M3(R2).” The EWG discussed the predictive value of photosafety testing in terms of possible risks to patients in phase 1 clinical trials and the potential recommendations that might result from a potential risk. There was also some discussion that an evaluation could be better collected as part of the safety assessment in a phase 1 trial. The FDA noted that phototoxicity was not thought to be a major observation in early clinical trials and thus did not warrant additional nonclinical testing. Ultimately, the EWG concluded that at the minimal, an assessment should be conducted. The EWG recognized that this was likely to become a topic of a future ICH guideline and for that reason decided not to incorporate more detailed recommendations.

13.2.12 Evaluation of Drug Metabolites

This topic needed to be addressed primarily because of the FDA guideline on the topic (Safety Testing of Drug Metabolites 2008). The FDA guideline states that a separate guideline would be coming out for anticancer drugs and this topic was being incorporated into the FDA draft guideline. When the ICH S9 topic was adopted by the Steering Committee, FDA participants argued that addressing drug metabolites then needed to be addressed by the ICH EWG.

In light of the FDA guidance on this topic, the EWG spent considerable time discussing this topic both before and after Step 2 and received extensive public comments. The Step 2 guideline stated that if the drug was positive in EFD or genotoxicity evaluations, then separate studies of the "disproportionate metabolite" might not be warranted. Several public comments stated that the intent was confusing; further, a definition of "disproportionate" could not be provided, and this could lead to different interpretations and disharmony. The EWG provided a more definitive conclusion in the Step 4 guideline, stating that a separate evaluation of metabolites identified in humans that may not have been qualified in animal studies was generally not warranted for patients with advanced cancer. The rationale for this approach is that for anticancer drugs, a maximum tolerated dose is usually studied nonclinically and clinically. For qualification purposes, the contribution of a metabolite to overall toxicity relative to the drug substance is generally expected to be low, and/or separate nonclinical studies with the metabolite alone are unlikely to provide additional value or change a clinical recommendation.

13.2.13 Evaluation of Combination of Pharmaceuticals

In the context of the guideline, combinations generally refer to coadministration of two or more pharmaceuticals. Some members of the EWG felt the combination toxicity data were needed, while others felt that it could be addressed by just automatically lowering combination doses in clinical studies. It was recognized that this later approach may not be optimal as it possibly could lead to under dosing of humans with cancer hoping for treatment, thus the recommendation to collect information of significant concern from combination pharmacology studies, even though these studies are not usually conducted according to good laboratory practice (GLP) regulations. The consensus opinion was that conducting an expanded pharmacology study should be the first step in understanding whether there is an increased risk of the combination compared to the individual compounds. This study would be particularly important for combinations in which at least one of the compounds was in early-stage development. To some parties, "early-stage development" generally referred to a pharmaceutical where the phase I study has not been completed (the human toxicity profile has not been characterized), although the EWG chose not to specify phase I in order to allow flexibility. This EWG affirmed that this study, as is typical for many

pharmacology studies, need not be GLP compliant. Of critical interest is whether there is significant change in severe toxicity that can be detected in combination pharmacology studies as it was recognized that the sensitivity of detecting small changes in toxicity in the expanded pharmacology studies was limited.

It should be noted that the phrase "specific cause for concern" is not found in the Step 4 document. The EWG noted that this phrase is somewhat vague and it would be more useful to provide clarity around what constituted concern. In the case of studying combinations, for example, concern generally refers to studying pharmaceuticals in which one compound of the combination is still early in development. In this case, a remedy is the expanded pharmacology study.

13.2.14 Flexibility in Qualification of Impurities

For potentially genotoxic impurities in genotoxic drug substances, it makes little sense to follow a threshold of toxicological concern approach to qualification. In addition, the threshold for toxicological concern approach addresses lifetime risk of cancer, and this is not considered appropriate for patients with advanced cancer. This would be particularly true if the genotoxic impurities arise late in development (e.g., as the commercial process becomes finalized) and the pharmaceutical has demonstrated a known survival advantage. Thus, the approach as outlined in the ICH guidelines Q3A and Q3B for determining levels for qualification may be more appropriate, although it should be noted that the EWG did not provide specific recommendations.

13.2.15 Examples of Treatment Schedules for Anticancer Pharmaceuticals to Support Initial Clinical Trials (Table 13.1)

For many participants, this table was thought to be one of the most useful parts of the document. This was also reflected in comments to the Step 2 document. The examples provided were not meant to take the place of rationale scientific judgment but to serve as a guide. The rationale for a single dose supporting a once-every-3–4-week schedule is that the experience is that this was the schedule for the traditional cytotoxic drugs that suppressed bone marrow, and full recovery took approximately 3 weeks. Otherwise, the rationale is that animals should be exposed to several doses of a compound at least but close as possible to the clinical schedule. It was recognized that the proposed clinical schedule may change during development so that some flexibility was needed to avoid repeating animal studies and delaying clinical trials. However, the EWG recognized that it was the sponsor's obligation to provide some justification that any proposed schedule would be safe for patients.

Table 13.2 Expected reduction in animal use by adoption of ICH S9

Study	Before S9	After S9	Effect
General toxicology	6 months' duration concurrent with drug development	1 month sufficient to initiate development; 3 months sufficient to support pivotal clinical trials	Long-term toxicology studies are not needed early in development, eliminating the need for many 3- and 6-month studies
General toxicology—change in clinical schedule	Need to conduct additional nonclinical studies (2 species?) to support a change in clinical schedule	Evaluate nonclinical and clinical data (if available) to see if available data is sufficient to support the proposed clinical schedule; if not, then a study in one species is usually sufficient	Reduction in the default requirement that studies in 2 species (rodent and non rodent) would normally be needed. Reduction in the use of nonrodents
General toxicology—recovery	Usually included in every study in 2 or more dose groups	Need based on scientific justification	Possible reduction in use of nonrodents; full effect remains to be determined
General toxicology—acute studies	Usually needed in every application	Not generally needed	Reduction in use of rodents
"Cytotoxic" drugs	2 species for initiating clinical development	1 species to initiate clinical development	Eliminated the need for the study in nonrodent to initiated clinical trials for "cytotoxic" drugs
Reproduction toxicology	Embryofetal studies needed in 2 species; studies needed fertility and pre- and postnatal development	Embryofetal study in one species only if positive	Eliminated the need for a second positive embryofetal toxicity study; no requirement for the fertility and pre- and postnatal development studies
Safety pharmacology	Follow ICH S7A and B	Cardiovascular safety and other endpoints could be incorporated into general toxicology studies	Minimized the need for stand-alone safety pharmacology studies
Photosafety testing	In vitro studies needed with possible in vivo follow-up	An assessment needed	No specific requirement for in vivo follow-up

13.2.16 The Role of 3Rs

Throughout the course of developing the ICH S9 guideline, the 3Rs, reduction, refinement, and replacement of animal testing had been kept in focus by the EWG to ensure that the goals outlined in the concept paper were realized. In this, the guideline was successful, eliminating or delaying some animal testing (Table 13.2). For example, recovery groups may not be needed if adequate scientific justification can be provided; acute toxicity studies are not needed; general toxicology studies are limited to 1 month to support initial clinical development, and 3 months should be sufficient to support phase III and in most cases the marketing application; and reproduction toxicology studies can be provided at marketing and are limited to studying embryofetal development.

13.3 Summary

The S9 EWG met between October 2007 and November 2009 and in that time developed a consensus as to what to include and not include in the tripartite guideline and incorporated public comments into the document. In reaching consensus, the EWG kept three principles in focus: patient safety; harmonizing requirements; and reduction, refinement, and replacement of animal use. The EWG met its timelines as set out in the business plan, and the guidance represents a significant step forward in harmonization. However, during the discussion and since publication, the guidance has raised several topics that may require additional discussion, including: what scientific data are needed to justify inclusion or noninclusion of recovery groups into a toxicology study; the number of dose levels to include in a nonrodent study; a more complete discussion surrounding the start dose of biologics, including antibody drug conjugates; and a more robust discussion of what would constitute an appropriate photosafety evaluation. Inclusion of this latter topic has resulted in some confusion as to what is needed. An interesting topic that could be included if the guidance were being written today is the development of dried blood spots for pharmacokinetic and toxicokinetic evaluation, as this could have the potential to reduce the need for satellite groups in rodent studies. The translation of the animal data into a clinical start dose still remains a challenge, particularly for biologics. In addition, while the EWG did have some discussion regarding the circumstances where a carcinogenic evaluation might be necessary, as this topic is addressed for anticancer drugs in ICH S1A, this topic did not receive as complete a discussion as warranted. As time progresses and more investigators gain experience with the guidance, other topics requiring clarification are sure to become evident. However, as one member of the EWG pointed out toward the end of the discussions, the EWG should not let a perfect guidance delay the good.

A legitimate question would be what effect the S9 guidance has had so far. From the regulatory perspective, by far the biggest impact of the guidance is anticipated

to be the reduction in the length of studies, from 6–9 months to 13 weeks to support product registration. One of the 10 new anticancer drugs approved since the beginning of 2011 was approved with 3-month-only studies for rodents and nonrodents. The toxicology studies for this drug were conducted in 2009. Of the remaining drug approvals, most of the toxicology studies for these programs date from the mid-2000s or before. Because of the lead time needed to conduct studies for development programs prior to filing for registration, it is premature to conclude that the reduction in the duration of the toxicology studies has made a major impact. However, sponsors are including questions relating to ICH S9 in meetings with FDA, including whether toxicology studies of 3 months duration would be sufficient to support further clinical development for drugs for patients with less advanced disease. Second, there has been some movement to 13-week-only studies for nonrodents to initiate clinical development of biologics, but this remains rare. In the past, 13-week study was often provided for initiating clinical development of a biologic to be studied in patients with cancer, perhaps due to an interpretation of ICH S6. The S9 guidance clearly states that studies of much shorter duration are acceptable for this purpose. The FDA has seen some toxicology studies without recovery groups, but most studies submitted still contain these add-on groups. Finally, the discussion in the guidance on combination of products provided much-needed clarity to this topic, especially with the growing clinical interest in trials using multiple drug combinations. Basically, the FDA is not seeing combination toxicology studies but well-designed pharmacology studies if needed, demonstrating the success of the guidance. Taken together, it is clear that another major impact of the guidance has been a reduction in animal use, a trend that is likely to continue.

The ICH Steering Committee signed off on the ICH S9 *Step 4* guideline in St. Louis following the recommendation of the ICH S9 EWG, after the EWG addressed the comments received after public consultation. Having produced a Step 4 harmonized guideline, the EWG accomplished its primary goal, but the task of implementation remains, a task made easier if the scientific rationale behind the recommendations is transparent.

Acknowledgments The authors wish to thank and acknowledge the helpful discussions, hard work, and contributions of the members of the ICH S9 EWG in the development of the guidance. These members did not participate in the writing of this article: Mikael Andersson, Marco Bughera, Osamu Fueki, James Green, Shuichi Kai, Daniel Lapadula, Tatiana Lejen, Helen Mao, Dai Nakae, Takahiro Nakazawa, Chihiro Nishimura, Anne Pilaro, Vijayapal Reddy, Fumio Sagami, Hermann Schweinfurth, Mercedes Serabian, Beat Schmid, and Ulla Wändel-Liminga.

References

DeGeorge JJ, Ahn CH, Andrews PA, Brower ME, Giorgio DW, Goheer MA, Lee-Ham DY, McGuinn WD, Schmidt W, Sun CJ, Tripathi SC (1998) Regulatory considerations for preclinical development of anticancer drugs. Cancer Chemother Pharmacol 41(3):173–185

EMA (1998) Committee for Proprietary Medicinal Products: Note for guideline on the pre-clinical evaluation of anticancer medicinal products; CPMP/SWP/997/96. EMA, London

EORTC New Drug Development Committee (1985) EORTC guidelines for Phase I trials with single agents in adults. Eur J Cancer Clin Oncol 21:1005–1007

Final Concept Paper (2007) S9: Preclinical guideline on oncology therapeutic development. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Safety/S9/Concept_papers/S9_Concept_Paper.pdf. Accessed 21 Mar 2012

ICH S5 (R2) (2000) Detection of toxicity to reproduction for medicinal products & toxicity to male fertility. http://www.ich.org/products/guidelines/safety/article/safety-guidelines.html.

ICH S6(R1) (2011) Preclinical safety evaluation of biotechnology-derived pharmaceuticals. http://www.ich.org/products/guidelines.html. Accessed 21 Mar 2012

ICH S5 (R2) (2000) Detection of toxicity to reproduction for medicinal products & toxicity to male fertility http://www.ich.org/products/guidelines/safety/article/safety-guidelines.html

Lowe MC, Davis RD (1987) The current toxicology protocol of the National Cancer Institute. In: Hellmann K, Carter SK (eds) Fundamentals of cancer chemotherapy. McGraw-Hill, New York

ODAC (2006): http://www.fda.gov/ohrms/dockets/ac/06/minutes/2006-4203M1.pdf. Accessed 12 July 2011

Prieur DJ, Young DM, Davis RD, Cooney DA, Homan ER, Dixon RL, Guarino AM (1973) Procedures for preclinical toxicologic evaluation of chemotherapeutic agents: protocols of the laboratory of toxicology. Cancer Chemother Rep Part 3 3(4):1–30

Safety Testing of Drug Metabolites (2008) http://www.fda.gov/downloads/Drugs/GuidelineComplianceRegulatoryInformation/Guidelines/ucm079266.pdf. Accessed 12 July 2011

Chapter 14
Non-clinical Safety Studies for the Conduct of Human Clinical Trials for Pharmaceuticals: ICH M3 and M3(R2)

Per Sjöberg and David R. Jones

Abstract The extent of non-clinical safety testing to support clinical trials at different stages of development differed greatly between the EU, Japan and the USA prior to the adoption of the original ICH M3 guidance in 1997. The guideline achieved some notable harmonizations, but there was still significant disharmony, especially around the duration of dosing for non-rodents and the timing and extent of reproductive toxicology studies to support trials in women of childbearing potential. The inability to harmonise on these particular issues led to a reluctant acceptance of finalising the M3 guidance.

In 2006, a revision of ICH M3 commenced with an aim to remove the un-harmonised components. Although the M3 guideline is essentially concerned with the timing of non-clinical studies in relation to clinical development, further topics were also introduced by the Expert Working Group during the discussions. The ICH M3(R2) document was signed off by the regulators in June 2009. While the 2000 version of the guideline had 6 pages of text, the revision had 27. All the objectives had been largely met and with only one minor difference still in place.

14.1 Introductory Comments

The extent of non-clinical safety testing to support clinical trials at different stages of development differed largely between the EU, Japan and the USA prior to the adoption of the ICH M3 guidance in 1997. These regional differences had been

P. Sjöberg (✉)
Eureda KB, Uppsala Science Park, Uppsala, Sweden
e-mail: per.sjoberg@eureda.com

D.R. Jones
Medicines and Healthcare products Regulatory Agency (MHRA),
Victoria, London, UK

J.W. van der Laan and J.J. DeGeorge (eds.), *Global Approach in Safety Testing*, AAPS Advances in the Pharmaceutical Sciences Series 5, DOI 10.1007/978-1-4614-5950-7_14,

highlighted and discussed at one of the safety workshops held at the First International Conference of Harmonization in Brussels in October, 1991 (Scales 1991). Although regional differences were apparent to international pharmaceutical companies making clinical trial submission in several regions, the exact differences were difficult to pinpoint as there was no clear guidance published on the timing of non-clinical safety studies in relation to stage of clinical development in either Japan or the USA, and in the EU community, there was only a draft guidance available with no formal status.

Although the ICH process early identified the need for the establishment of clear and internationally harmonised recommendations on the extent of non-clinical safety studies needed to support clinical trials of different phases, it was not until 1994 the "M3 project" was formally adopted as an ICH topic and an Expert Working Group (EWG) was formed. This delay in the initiation of the more formal work with the M3 guidance was logical considering that several important ICH safety topics related to the scope of the M3 guidance were at the initial stage of development, and thus the timing question which could not be addressed prior to these safety topics had become more mature in terms of regional harmonization. For example, until there had been a position on the type of genotoxicity studies that would be acceptable to support a marketing authorisation (ICH S2B), it would be impossible to adopt a clear and understandable guidance what type of data would normally be needed to support clinical trials of different stages/durations. Other examples of specific ICH safety topics worked on were the guidance on Duration of Non-Rodent Toxicity Testing (S4), Reproduction Toxicity (S3) and Carcinogenicity [particularly S1A (Need for Carcinogenicity Testing of Pharmaceuticals)].

The object of the ICH Steering Committee to initiate work on a regulatory guidance on the timing of non-clinical safety studies was obviously to create international harmonization, i.e. harmonization of the recommendations/requirements from the regulatory authorities in EU, Japan and the USA. However, the lack of clear regional positions on many timing aspects prior to the initiation of this work pushed the regional authority groups and the regional industry groups to formulate updated positions on these aspects. This meant that the discussions and negotiations were made from an essentially equal playing field which fostered an open and constructive dialogue. It is perhaps unknown to many that the M3 guidance that was adopted by the three regional authorities in 1997 had, from a formal point of view, a weak position in the EU. At this time, there was no European Clinical Trial Legislation (the EU Clinical Trials Directive did not come into force until May 2004), and thus the guidance was not binding to the EU member states even though the CPMP had adopted the guidance. However, in view of the lack of guidance relating to the extent of non-clinical safety testing to support clinical trials and divergent regulatory scrutiny of clinical trial applications within the EU member states, the development of the ICH M3 guidance was perhaps of particular significance for EU in that it both catalysed necessary harmonization within EU and provided an important basic element for the forthcoming EU Clinical Trial Directive.

The EU, Japanese and US pharmaceutical companies were understandably eager proponents of the M3 guidance. Since a large proportion of the companies were working on the international market, they were keen to obtain harmonization on this important subject particularly as inconsistent regulatory request for non-clinical safety studies to support either national or multinational clinical trials slowed clinical development and thereby incurred additional costs. As indicated above, the regulatory participants of the EWG from the three regions were also eager to work towards a harmonised guidance in part because of the awareness that the lack of clear regional guidance was very unsatisfactory for drug developers and that without some type of harmonization of the timing issues, the achievements made in harmonising the technical standards for non-clinical safety testing (the other ICH topics) would not be fully appreciated.

14.2 Overall Content of the M3 Guidance

Once the ICH Steering Committee in 1994 approved a proposal that work should commence on a guidance addressing timing of safety studies in relation to clinical trials, the appointed Expert Working Group rapidly came to an agreement that the guidance should focus on the following principle areas of toxicity testing supporting clinical development:

- Safety pharmacology studies (effects on vital organ systems)
- Single- and repeat-dose toxicity studies
- Genotoxicity and carcinogenicity studies
- Toxicity to reproduction

Moreover, it was agreed that the principles of toxicokinetics needed a prominent place in the guidance since this was becoming more generally recognised as a fundamental part of non-clinical safety assessment. Other safety areas that were included without any controversy were local tolerance data and data to support clinical trials in paediatric populations.

14.3 Safety Pharmacology Studies

It is interesting to note that in the initial review of the timing issue presented to the participants of ICH I in Brussels in 1991 (Scales 1991), safety pharmacology was not included as one of the areas that needed to be addressed from a timing perspective. However, the request for an assessment of effects on vital functions such as cardiovascular, central nervous and respiratory system was soon incorporated in the guidance by the EWG and without much controversy. The wording “assessment of effects on vital functions” was carefully chosen to imply that such information could be

obtained in conjunction with single dose or more likely repeat-dose toxicity studies. This was also mentioned in the guidance.

The ICH M3 guidance from 1998 (ICH M3 1997) contains no reference to whether the non-clinical safety studies should be conducted according to GLP. Since this was already a requirement for all toxicity studies, it was argued by many EWG members (particularly those from the FDA and US Pharma) that this did not need to be included in the document. However, as there at the time was no specific guidance on safety pharmacology, studies conducted to assess the effect on vital organ system could, at least in some regions, be conducted without GLP compliance. In Europe, there was already a CPMP guidance clarifying the 91/507/EEC Directive regarding GLP and safety tests stating that "pharmacodynamic studies designed to test potential for adverse effects" must conform to GLP (CPMP III 3824 92 Rev). Thus, even though, as mentioned above, there was no harmonization of the legislation relating to applications and conduct of clinical trials within the EU, safety pharmacology data, when submitted for a marketing authorisation application, was expected to be derived from GLP-compliant studies. It should be noted that the ICH S7A guidance from 2001 (ICH S7A 2000) that addresses the specifics of safety pharmacology testing does address the GLP issue of safety pharmacology data. It is encouraging that this guidance gives some flexibility with regard to GLP compliance.

14.4 Single- and Repeat-Dose Toxicity Studies

The timing of single-dose toxicity studies was obviously of no controversy as such data should logically be available prior to first dose in humans. At the time when the work with M3 was initiated, international harmonization had already been achieved with regard to the number and of type of single-dose studies needed to support human clinical trials (Ohno 1991). Although several M3 EWG members likely felt that specific acute toxicity studies had limited value for human risk assessment, it was not possible to reopen an issue that was just recently harmonised and promoted as a major achievement of the ICH process. It is therefore of great satisfaction that the revised M3 document from 2009 (ICH M3(R2) 2009) abandons the request for specific GLP-compliant single-dose studies with two routes of administration and instead recommends that acute or single-dose toxicity information may be derived from dose-finding studies to support dose setting of repeat-dose studies.

One of the most difficult areas to harmonise during the entire ICH process has been the extent of repeat-dose toxicity data needed to support clinical trials of different stages and durations. This difficulty surfaced already at the beginning of the work with the M3 guidance as reported at the Third International Conference of Harmonization in Yokohama in 1995 (Hayashi 1995; Sjöberg 1995). Consensus could not be reached in the following areas:

(a) Duration of rodent toxicity studies to support single-dose and repeat-dose trials of up to 14 days (Japan requested 4-week rodent studies, while EU and USA accepted 2-week rodent studies)
(b) Duration of rodent and non-rodent studies to support phase III trials (EU requested longer study durations than Japan and USA)
(c) Duration of non-rodent studies needed to support clinical trials with a duration of more than 6 months (USA needed studies of 12-month duration, while EU was content with a 6-month study duration)

These regional differences were essentially based on what was considered a safe approach for clinical trial testing, i.e. no solid data was presented by either region to support its specific position. The Japanese argued that a 4-week rodent toxicity study was needed to assess the potential of the drug in question to interfere with male fertility and that such an assessment was needed for even the shortest clinical trial in humans. The EU regulators on the other hand argued that for confirmatory clinical trials, i.e. phase III trials, more solid toxicity data (longer term studies) was needed to establish the true toxicity profile and assess patient safety compared to the exploratory trial situation where homogeneity of patient population and patient numbers were different. Finally, the USFDA argued that in their experience, there was additional value of having non-rodent toxicity data from 9- or 12-month exposure duration as compared to the 6 advocated particularly by the EU. Owing to the failure of the EWG in reaching consensus on the maximum duration of non-rodent repeat-dose testing, a specific EWG was set up, and the ICH S4 topic was created in 1997, i.e. just prior to the finalisation of the M3 guidance. Based on an assessment of a limited data set of non-rodent repeat-dose toxicity studies that covered both 6- and 12-month exposure, the S4 EWG was also unable to come to a consensus on the maximal duration of repeat-dose studies (ICH S4 1998). Although the ICH Steering Committee was not pleased with the inability of the EWG to harmonise the timing and duration of repeat-dose toxicity studies, the EWG members, particularly those from the regulatory side, were less concerned with this inability. Their views were incorporated in the guidance, and it was felt that the differences did not overshadow the overall achievements in harmonising timing issues.

14.5 Genotoxicity and Carcinogenicity Studies

In co-operation with the EWG on genotoxicity, the M3 EWG rapidly came to a consensus that phase I clinical trials should be amply supported by the so-called ICH standard battery of in vitro studies while phase II trials should be supported by the complete set of ICH-compliant in vitro and in vivo studies (Mayahara 1995). In hindsight, one may question the total logic in this rigid separation between data request for phase I and those for phase II trials. The exposure duration may certainly

be longer in some phase I trials than in some phase II trials, and a recommendation based on exposure duration rather than what stage of development the drug is at would be a more logical approach considering that a genotoxic liability would likely be strongly linked to exposure duration and total dose. It should be noted that the updated M3 guidance (ICH M3(R2)) states that single-dose trials are generally supported by an assay for gene mutation, and thus this new guidance has adopted, from this perspective, a more logical approach to request for genotoxicity data. However, multiple-dose phase I studies, regardless of exposure duration, are seemingly still supported by the in vitro studies, while the complete set of in vitro and in vivo studies are needed prior to the conduct of phase II trials, again regardless of exposure duration.

The issue of timing of carcinogenicity studies was not primarily dealt with by the M3 EWG since a specific guidance relating to the "Need for Carcinogenicity Studies" (ICH S1A) was developed and adopted as a step 4 document by the ICH Steering Committee already in November 1995, i.e. shortly after the initiation of the work on the M3 guidance. In relation to the timing issue, the S1A guidance concluded that "when carcinogenicity studies are required they usually need to be completed before application for marketing approval", and unless there is specific concern, carcinogenicity data would not be needed prior to the conduct of large clinical trials. An example of where regulators subsequently have recommended carcinogenicity testing prior to large-scale clinical trials/patient treatment duration longer than 6 months is with the peroxisome proliferator-activated receptor (PPAR) agonists (CDER 2008; EMEA/341972/2006).

14.6 Toxicity to Reproduction

Similar to the situation for the timing of repeat-dose toxicity testing, the three regions, EU, Japan and USA, could not reach complete consensus of what toxicity to reproduction studies was needed to support the inclusion of men and women of childbearing potential in clinical trials of different development stage. Consensus could not be reached in the following two areas:

(a) Extent of repeat-dose data to make an assessment of potential to interfere with male fertility (in Japan, a 4-week repeat-dose toxicity study was considered essential to assess toxicity to the male reproductive system, whereas in EU and USA, 2-week toxicity studies were considered sufficient for an overall assessment of potential toxicity).
(b) Type of data needed to include women of childbearing potential, using highly effective birth control, in shorter-term clinical trials (assessment of female fertility and embryo–foetal development is needed in Japan, and embryo–foetal development studies are needed in EU, whereas in USA, women of childbearing potential could be included in "early, carefully monitored" clinical studies provided that "adequate precautions were taken to minimise risk").

It may be difficult to appreciate the original Japanese position that a 4-week toxicity study was needed to assess potential effect on male fertility even for a drug that is to be given for a single dose, but then one need to consider that in Japan, male fertility studies were needed to support inclusion of men in any clinical trial prior to the adoption of the M3 guidance in 1997. The Japanese should also be credited for performing experimental studies to support their new position that a 4-week rat toxicity study was sensitive to pick up potential effects on male reproductive organs (Takayama et al. 1995). In an additional collaborative study in Japan, rodent data were obtained supporting the position that 2-week toxicity studies in rats were as sufficient as 4-week studies to identify male reproductive toxicants (Sakai et al. 2000). The M3 guidance was therefore updated to include this new position of the MHW (ICH M3(R1)) and has been kept in the most updated version of ICH M3, i.e. ICH M3(R2).

The EWG discussions on the acceptability of including women of childbearing potential with adequate contraception in early clinical trials were fairly straightforward with no real attempt by either the EU or Japanese regulators to convince the USA that the more stringent position was correct. The EU regulators of the EWG had clear sympathy for the US position as their approach seemingly had been shown to work safely. When this position was forwarded to the CPMP, strong oppositions were given from a couple of the leading members, and there was thus no way a change in attitude could come about from EU regulators. The inability to harmonise on this particular issue and on the maximum duration of non-rodent toxicity studies was almost sufficient to stop the work on the M3 guidance by EU regulators not close to the actual work. When the overall benefits of all the harmonization that were achieved were enforced, there was a reluctant acceptance of finalising the M3 guidance. It is noteworthy that the Japanese and EU regulators have moved its position on this topic to the position USFDA was in 1998 (ICH M3(R2)).

14.7 ICH M3(R2)

At the ICH Steering Committee in early 2006, it was agreed that the ICH M3 guideline required further revision to try and achieve closer harmonization in non-clinical testing of pharmaceuticals. The issues to be discussed in the revision process were agreed and included the nature and timing of reproductive toxicity studies to support the conduct of different phases of clinical trials, the duration of repeated-dose toxicity studies to support the conduct of different phases of clinical trials, the duration of chronic toxicity studies in non-rodents, the requirement of the toxicity package to support first entry into human and the definition of the role of the M3 guideline in the development of biotechnology derived.

Although the M3 guideline is essentially concerned with the timing of non-clinical studies in relation to clinical development, further topics were introduced by the Expert Working Group during the discussions and included the removal of the need to keep single-dose toxicity studies as a fixed requirement prior to first human exposure.

In addition to regulatory and industry representatives for the three ICH regions, the working group also included observers from the European Free Trade Association (EFTA), Health Canada and the interested parties, the International Generic Pharmaceutical Alliance (IGPA) and the Biotechnology Industry Organisation (BIO).

The discussions surrounding the scope of the revised guideline centred on achieving a document that should facilitate the timely conduct of clinical trials, reduce the use of animals in accordance with the 3R (reduce/refine/replace) principles and reduce the use of other drug development resources.

The discussions surrounding whether single-dose toxicology studies were needed were supported by a publication on a European pharmaceutical company initiative challenging the regulatory requirement for acute toxicity studies in pharmaceutical drug development (Robinson et al. 2008) that followed on from work in 2007 by the UK's National Centre for the 3Rs (NC3Rs).

A considerable amount of time was taken to try and harmonise the nature and timing of reproductive toxicity studies to support the conduct of different phases of clinical trials. Industry associations from all three ICH regions provided vast databases on reproductive toxicology studies and publications on work to evaluate toxicity on male reproductive organs by 2-week repeated-dose toxicity studies in rats (Sakai et al. 2000) and on the evaluation of ovarian toxicity by repeated-dose and fertility studies in female rats (Sanbuissho et al. 2009). The final document almost achieved complete harmonization, with only one minor difference still in place. In the United States, assessment of embryo–foetal development could be deferred until before phase III for women of childbearing potential (WOCBP) using precautions to prevent pregnancy in clinical trials. In the EU and Japan, small numbers (about 150) of WOCBP could now be included in clinical trials of short duration, but definitive non-clinical developmental toxicity studies were still needed to be completed before exposure of large numbers. This represented a significant shift in opinion from the EU as previously most member states would not allow clinical trials in WOCBP without the results from reproductive toxicology studies.

The ICH M3(R2) document was signed off by the regulators in June 2009. The title had changed and was now "Guidance on the Non-Clinical Safety Studies for the Conduct of Human Clinical Trials and Marketing Authorisations for Pharmaceuticals". While the 2000 version of the guideline had 6 pages of text, the revision had 27. All the objectives had been met, and, in addition, the new guideline had sections on High Dose Selection for General Toxicity Studies, i.e. Safety Margins, Metabolites in Safety Testing, Estimation of the First Dose in Humans, Exploratory Clinical Trials, Immunotoxicology, Phototoxicity, Non-Clinical Abuse Liability and Combination Drug Non-Clinical Testing. These sections were added as it was agreed that there was a lack of regulatory guidance in these areas. The new guideline also included reference to juvenile animal toxicology studies to support clinical trials in paediatric populations.

The new section on Exploratory Clinical Trials was extremely interesting. It was recognised that in some cases earlier access to human data can provide improved insight into human physiology/pharmacology, knowledge of drug candidate characteristics and therapeutic target relevance to disease. The USFDA had

already published guidance on this subject, and the EU's Safety Working Party had published a concept paper and was drafting their guideline. Once it was decided to include a section within ICH M3, the EU stopped their independent work. While ICH M3 is essentially a timing document, i.e. it advises when studies are required during clinical development; this new section included definitive advice on the type and design of studies.

Exploratory clinical studies for the purpose of this guidance were considered to be those intended to be conducted early in phase I, involve limited human exposure, have no therapeutic intent and were not intended to examine clinical tolerability. Recommended starting doses and maximal doses for the five approaches were also included. Five study types were included, two micro-dose designs, one single-dose design and two repeated-dose designs, one based on the FDA guidance and one based on the "EU approach".

Another notable achievement was the addition of the new section "High Dose Selection for General Toxicology Studies". Generally, in toxicity studies, effects that are potentially clinically relevant can be adequately characterised using doses up to the maximum tolerated dose (MTD). Other equally appropriate limiting doses include those that achieve large exposure multiples or saturation of exposure or use the maximum feasible dose (MFD). This section was added to prevent the use of doses in animals that would not add value to predicting clinical safety, and the recommendations were consistent with those for ICH reproduction and carcinogenicity study designs that already had defined limit doses and/or exposures.

While still in its early phases of the implementation, and even though the document had been released for public comments on two occasions, the complexity of the guidance, its broader scope and the numerous changes in recommendations from the ICH M3(R1) guidance generated questions that could have impacted on its successful implementation.

Several of these questions and issues were considered to be outside the scope of the guideline, while others were addressed by question and answer (Q&A) documents that were released in 2011 and 2012. The issues covered in the Q&A documents were limit doses, exploratory clinical trials, reversibility of findings, metabolite testing, juvenile animal toxicology studies, reproductive toxicology studies and safety pharmacology.

14.8 Concluding Remarks

The first international guidance document addressing the aspect of timing on non-clinical safety studies in relation to clinical trials should in hindsight be viewed as a success for regulators and pharmaceutical companies alike although complete harmonization could not be reached on all timing issues. It was particularly useful for the member states of the EU that from 2004 were obliged to follow the EU Clinical Trials Directive (2001/20/EC). The M3 guidance document was obviously an important component of the implementation of this guidance.

The inability of the three regions to harmonise on all areas of non-clinical safety testing to support clinical trials should be judged from the fact that an exact value of non-clinical safety testing in predicting human safety cannot be given, and therefore there will always be an element of personal/regulatory agency judgment in defining what type of studies is necessary to safeguard patients in a particular clinical trial situation. The significant expansion of the ICH M3(R2) document and the subsequent issuing of explanatory Q&A documents are interesting and are probably further reflections on this point.

Lastly, it should be emphasised that the ICH process has overall made very significant contributions in underpinning the scientific basis for various standards and recommendations by encouraging retrospective analysis of non-clinical safety data and the initiation of prospective studies. Many of these contributions have had a direct impact on recommendations made in the M3 document. If such efforts continue, the M3 document will maintain its status as one of the most important non-clinical regulatory documents.

References

2001/20/EC. Directive 2001/20/EC of the European Parliament and of the Council of 4 April 2001on the approximation of the laws, regulations and administrative provisions of the Member States relating to the implementation of good clinical practice in the conduct of clinical trials on medicinal products for human use. Official Journal of the European Communities 1.5.2001. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2001:121:0034:0044:en:PDF

CDER (2008) Guidance for industry. Diabetes Mellitus: developing drugs and therapeutic biologics for treatment and prevention. Center for Drug Evaluation and Research, Feb 2008. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM071624.pdf?utm_campaign=Google2&utm_source=fdaSearch&utm_medium=website&utm_term=PPARguidance2008&utm_content=2

CPMP III/3824/92 Rev.1. Committee for Proprietary Medicinal Products. Applicability of good laboratory practice. Commission of the European Communities

EMEA/341972/2006. CHMP SWP reflection paper on PPAR's (Peroxisome Proliferator Activated Receptors), Sept 2006. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/11/WC500015479.pdf

Hayashi (1995) Duration of repeated dose studies to support first entry into humans. In: D'Arcy and Harron (eds) Proceedings of the third international conference on harmonization, Yokohama, pp 342–343

ICH M3 (1997) Nonclinical safety testing for the conduct of human for pharmaceuticals. ICH, July 1997. http://ocw.jhsph.edu/courses/drugdevelopment/PDFs/FDA_Guidance_on_NonCLinical_Safety_Studeis.pdf

ICH M3(R2) (2009) Guidance on nonclinical safety studies for the conduct of human clinical trials of and marketing authorization for pharmaceuticals. 11 June 2009. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Multidisciplinary/M3_R2/Step4/M3_R2__Guideline.pdf

ICH S4 (1998) Duration of chronic toxicity testing in animals (rodents and non-rodents). 2 Sept 1998. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Safety/S4/Step4/S4_Guideline.pdf

ICH S7A (2000) Safety pharmacology studies for human pharmaceuticals. 8 Nov 2000. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Safety/S7A/Step4/S7A_Guideline.pdf

Mayahara (1995) Progress of the ICH guideline. In: D'Arcy and Harron (eds) Proceedings of the third international conference on harmonization, Yokohama, pp 339–342

National Centre for the Replacement, Refinement and Reduction of Animals in Research (2007) Challenging requirements for acute toxicity studies: Workshop report

Ohno (1991) Toxicity testing: regulatory perspectives. In: D'Arcy and Harron (eds) Proceedings of the first international conference on harmonization, Brussels, pp 186–188

Robinson S, Delongeas JL, Donald E, Dreher D, Festag M, Kervyn S, Lampo A, Nahas K, Nogues V, Ockert D, Quinn K, Old S, Pickersgill N, Somers K, Stark C, Stei P, Waterson L, Chapman K (2008) A European pharmaceutical company initiative challenging the regulatory requirement for acute toxicity studies in pharmaceutical drug development. Regul Toxicol Pharmacol 50:345–352

Sakai et al (2000) Collaborative work to evaluate toxicity in male reproductive organs by repeated dose studies in rats – overview of the studies. J Toxicol Sci 25:1–21

Sanbuissho A, Yoshida M, Hisada S, Sagami F, Kudo S, Kumazawa T, Ube M, Komatsu S, Ohno Y (2009) Collaborative work on evaluation of ovarian toxicity by repeated-dose and fertility studies in female rats. J Toxicol Sci 34:SP1–SP22

Scales (1992) Timing of toxicity studies vs conduct of clinical trials – existing requirements. In: D'Arcy and Harron (eds) Proceedings of the first international conference on harmonization, Brussels, pp 322–329

Sjöberg (1995) Data to support the extension of clinical trials. In: D'Arcy and Harron (eds) Proceedings of the third international conference on harmonization, Yokohama, pp 343–345

Takayama et al (1995) A collaborative study in Japan on optimal treatment period and parameters for detection of male fertility disorders induced by drugs in rats. J Am Coll Toxicol 14:266–292

Biography

Joseph J. DeGeorge has a Ph.D. in Pharmacology from SUNY Upstate Medical Center, Syracuse, NY; was a Postdoctoral Fellow at UNC, Chapel Hill, NC, and Burroughs Wellcome Research Institute; and was a Senior Staff Fellow at the NIH/NIA in the Laboratory of Neuroscience. He has been actively engaged in the Safety and Multidisciplinary topics of the ICH as an Expert, Topic Rapporteur, and Safety Lead on numerous topics since the early 1990s. He is still actively engaged in ICH as of 2012. He has served these roles originally as a Food and Drug Administration representative and, following a brief hiatus, returned to the ICH as a PhRMA representative. He began his participation in the ICH process while serving as a Pharmacology and Toxicology reviewer in the Neuropharmacology Division, CDER/FDA, and took on increasing responsibility as his career progressed, culminating in the role of FDA ICH Safety Lead and Associate Director for Pharmacology and Toxicology for the Office of New Drugs, CDER/FDA. He joined the pharmaceutical industry in 2002 and has been Vice President of Global Safety Assessment at Merck & Co. since 2004. He also serves as the PhRMA Lead for ICH Safety topics, participating in the development of numerous new and revised ICH Safety Guideline with a goal of enabling the safe and efficient development of new medicinal products.

Dr. Jan Willem van der Laan was from 1990 to 2007 heading the Section Pharmacology and Toxicology Assessment at the National Institute for Public Health and the Environment (RIVM, Bilthoven). In this function, he was responsible for the advice on non-clinical safety aspects for the Netherlands "College", the Medicines Evaluation Board. Since September 2007, he is still senior assessor in pharmacology and toxicology. He was from 2002 to 2010 project leader of the Teratology Information Service at the institute. He moved in 2012 to the Medicines Evaluation Board located in Utrecht, but his role as senior assessor has not been changed. On behalf of this Board, he is a member and vice-chair of the Safety Working Party (SWP) of the CHMP. His contributions to the International Conference on Harmonisation started in early 1992 when he became a member of the Expert

J.W. van der Laan and J.J. DeGeorge (eds.), *Global Approach in Safety Testing*, AAPS Advances in the Pharmaceutical Sciences Series 5, DOI 10.1007/978-1-4614-5950-7,

Working Group on Carcinogenicity Testing. Later on, he became EU rapporteur for another topic in the safety area, i.e. immunotoxicology. In 2011, he finished as the EU rapporteur for the Addendum for the Preclinical Testing of Biotechnology-Derived Proteins (ICH S6). In 2012, he is again involved in the new start of a topic on carcinogenicity.

Index

J.W. van der Laan and J.J. DeGeorge (eds.), *Global Approach in Safety Testing*, AAPS Advances in the Pharmaceutical Sciences Series 5, DOI 10.1007/978-1-4614-5950-7,