

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

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 assess 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 SIC 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/m² 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/m² 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/m² 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 SIC 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 used 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/m² 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 SIC(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 SIC guidance. Thus inclusion of transgenic animal dose selection could not even be contemplated at the time of the original guidance. However, revisions of SIC 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 SIC 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 than 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	7	8	1	2	83% (15/18)
XPA ^{-/-} /p53 ^{+/-}					

^aCompounds 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)

^bGenotoxic compounds only

^cResponses 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 be 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 and 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 nongenotoxic 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 pre-clinical 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 (Lempiäinen 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, Lempiäinen 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, Krayner JH, Martin P, Descotes J (2010) Critical review of pre-clinical 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 artherosclerosis 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-Stephen 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, Cunningham 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
- Lempiäinen 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 Experimentia* 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
- Walitzka 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