

Improving Prediction of Carcinogenicity to Reduce, Refine, and Replace the Use of Experimental Animals

Todd Bourcier,* Tim McGovern, Lidiya Stavitskaya, Naomi Kruhlak, and David Jacobson-Kram[†]

Cancer risk assessment of new pharmaceuticals is crucial to protect public health. However, clinical trials lack the duration needed to clearly detect drug-related tumor emergence, and biomarkers suggestive of increased cancer risk from a drug typically are not measured in clinical trials. Therefore, the carcinogenic potential of a new pharmaceutical is extrapolated predominately based on 2-y bioassays in rats and mice. A key drawback to this practice is that the results are frequently positive for tumors and can be irrelevant to human cancer risk for reasons such as dose, mode of action, and species specificity. Alternative approaches typically strive to reduce, refine, and replace rodents in carcinogenicity assessments by leveraging findings in short-term studies, both in silico and in vivo, to predict the likely tumor outcome in rodents or, more broadly, to identify a cancer risk to patients. Given the complexities of carcinogenesis and the perceived impracticality of assessing risk in the course of clinical trials, studies conducted in animals will likely remain the standard by which potential cancer risks are characterized for new pharmaceuticals in the immediate foreseeable future. However, a weight-of-evidence evaluation based on short-term toxicologic, in silico, and pharmacologic data is a promising approach to identify with reasonable certainty those pharmaceuticals that present a likely cancer risk in humans and, conversely, those that do not present a human cancer risk.

Abbreviations: DRA, drug regulatory agency; CDER, Center for Drug Evaluation and Research; FDA, US Food and Drug Administration; ICH, International Conference on Harmonization; WOE, weight of evidence.

The United States Food and Drug Administration (FDA), like its international drug-regulatory counterparts, is dedicated to approving safe and effective drugs for human use. In addition, the FDA is committed to the 3Rs, by reducing, refining, and replacing the use of animals in drug development. Because drugs are designed to have biologic effects and because they are often taken for prolonged periods of time, potential toxicities are concerns. Although most potential toxicities are discovered in the course of clinical trials, some endpoints such as carcinogenicity, mutagenicity, and teratogenicity can only be assessed in nonclinical studies in light of ethical and practical considerations. Here we review current regulatory practices and challenges to assessing carcinogenic risk, and we explore emerging approaches all within the framework of the 3Rs.

The long latency period for most human cancers precludes the assessment of potential carcinogenicity in the course of clinical trials. Even retrospective epidemiologic studies lack sensitivity because of high backgrounds and multiple drug exposures. As a result, the 2-y rodent bioassay has become the ‘gold standard’ for assessing potential carcinogenicity of new drugs as well as other products.

With very few exceptions, rodent bioassays detect all known human carcinogens. Nevertheless, few would deny that the rodent bioassay has flaws. The studies are relatively insensitive because of the small numbers of animals used compared with the number of people exposed. The studies are lengthy—after accounting for the time needed to complete dose-range

finding studies and preparation of the final study report, total turnaround times can be over 3 y. Although the costs of carcinogenicity assessments account for only a small fraction of the overall cost of drug development, the costs of carcinogenicity assessments can run into millions of dollars. Doses exceeding those likely to be used clinically are often tested; the high dose in a cancer bioassay is often a maximally tolerated dose. Exposing animals to maximally tolerated doses can alter biologic processes that are not relevant at clinical exposures and can produce artifacts, necessitating careful interpretation of positive tumor results. In addition, carcinogenicity studies frequently are criticized because rodents are perceived as too biologically different from humans and therefore poor models for assessing cancer risks. Finally, rodent bioassays can result in drug-induced tumors that arise by mechanisms of questionable relevance to human risk.³⁰

Previous Attempts to Substitute In Vitro and Short-Term Tests for Rodent Carcinogenicity Studies

In the mid-1970s, Professor Bruce Ames made the important observation that chemicals known to be carcinogens often induced mutations in bacterial cells.² This observation led to the development of the *Salmonella typhimurium* reverse-mutation assay, often referred to as the Ames assay. Other endpoints of genotoxicity were known as well, including chromosomal aberrations in metaphase cells, micronuclei, sister chromatid exchanges and unscheduled DNA synthesis. At that time it was thought that a battery of in vitro assays could successfully identify potential carcinogens and thereby substitute for 2-y rodent bioassays. However, a landmark study published in

Received: 27 Feb 2014. Revision requested: 24 Mar 2014. Accepted: 08 May 2014.
US Food and Drug Administration, Center for Drug Evaluation and Research, Silver Spring, Maryland.

[†]ToxRox Consulting, 5910 Chesterbrook Road, McLean, VA 22101

*Corresponding author. Email: todd.bourcier@fda.hhs.gov

1987³³ showed that although clearly genotoxic chemicals were often carcinogens, the reverse was not true—there are many nongenotoxic carcinogens. Therefore a battery of assays that only identifies genotoxicity could not substitute for the 2-y rodent bioassay.

Over the years other *in vitro* assays have been developed with the goal of supplanting the need for lifetime rodent bioassays. For example, the Syrian hamster embryo transformation assay was touted to have this potential, especially a variation on the assay that used a pH that was lower than physiologic pH.²³ The popularity of that assay waxed and waned, but the test was finally rejected as not being useful in drug development. Mechanisms by which drugs induce cancer often relate to exaggerated pharmacology, hormonal imbalance, and immune suppression. These mechanisms are unlikely to be replicated in a Syrian hamster embryo assay.

Progress on In Silico Prediction of Carcinogenicity

The use of *in silico* models has become increasingly important for risk assessment, because these models offer early insight into potential safety issues for a new pharmaceutical agent. Quantitative structure–activity relationship (Q)SAR models can provide a rapid and cost-effective assessment of the toxicity of a molecule based solely on its chemical structure.²² (Q)SAR models can be used to make predictions for a variety of chemical types, but their performance varies depending on attributes of their training set including quality of data, size of training set, range of structural diversity, ratio of active to inactive chemicals, and mechanistic complexity of the endpoint.²⁷ Carcinogenicity is one of the most difficult toxicologic endpoints to predict, given its mechanistic complexity and high variability of long-term study data. Therefore it is particularly important to understand the value and limitations of available (Q)SAR models before selecting an appropriate model for a specific analysis.

The success and quality of any (Q)SAR model greatly depends on the data that are used to link chemical structures to biologic outcomes. FDA's Center for Drug Evaluation and Research (CDER) has expended considerable effort to construct and maintain a database of 2-y rodent carcinogenicity findings from drugs and industrial chemicals that have been used for the development of (Q)SAR models as well as for internal structural analog searching. The database was first described in 1998²⁵ and, since then, has been continually updated²² with new studies for orally dosed pharmaceuticals obtained through CDER submissions and other sources such as the National Toxicology Program. In addition, the database has been expanded to include organ-specific findings and supporting citations in a standardized and searchable format, and it is being made available internally to CDER reviewers. The database now comprises 1534 unique chemical substances but, due to the high cost of performing these studies, the rate of new incoming data is relatively low, with the addition of approximately 50 new studies each year. Although the FDA–CDER database represents one of the most comprehensive and high-quality data sets available for rodent carcinogenicity, the low availability of new data compared with other types of toxicology study findings still presents a challenge to the development of predictive models for such a mechanistically complex endpoint.

Two basic types of structure-based modeling approaches have been used at FDA–CDER to assess the carcinogenic potential of chemicals: (1) statistical-based modeling[(Q)SAR], and (2) expert rule-based modeling (SAR). Statistical (Q)SAR models identify a mathematical correlation between molecular

features and activities by using a machine-learning approach, whereas an expert rule-based approach uses human-derived, and often mechanistically defined, structural alerts to assess biologic activity. Although expert systems typically provide greater interpretability, they are much more time-consuming to develop and may lack clarity in the absence of a structural alert. In contrast, statistical methods allow the identification of correlations across large datasets that might otherwise be impossible to detect through manual inspection; however, they can be prone to identifying correlations that are statistically significant but biologically meaningless.²⁷ To benefit from the strengths of each methodology, a combination of both has been used for predicting carcinogenicity based on the FDA–CDER carcinogenicity database and, in conjunction with expert knowledge, is the accepted overall approach for the use of (Q)SAR to qualify drug impurities for mutagenic potential under the newly released M7 guideline from the International Conference on Harmonization (ICH).²⁰

In addition to the general categories of statistical- and expert-rule-based systems, (Q)SAR models can further be described as local or global, reflecting the structural diversity of compounds from which they are derived and for which they are applicable. Local models are constructed using highly structurally similar chemicals and have the advantage of providing more accurate predictions for a specific class, but with the disadvantage of having only a narrow region of chemical space in which they are applicable. Examples of local models for carcinogenicity include those for chemical classes such as aromatic amines^{4,12,13} and N-nitrosamines.³ In contrast, global models are constructed using large and structurally diverse sets of chemicals and can be used to predict toxicity across different structural classes and chemical types. Such models show greater utility in a regulatory setting, which can require assessment across broad chemical structural space covering both drug-like and industrial chemicals. Recently, the FDA–CDER carcinogenicity database was used to construct new global (Q)SAR models trained on non-proprietary data and externally validated by using both proprietary and non-proprietary data for drugs and industrial chemicals, in part from FDA archives. Overall accuracy of these models ranged from 61% to 71%, with a focus on negative predictivity (64% to 84%) and sensitivity (60% to 81%) to protect patient safety by minimizing the risk of a false negative prediction.³²

The use of (Q)SAR models for carcinogenicity may increase as they become more widely accepted as a replacement for conventional toxicology testing of pharmaceuticals. Indeed, the relatively low cost in time and resources and the consistency and transparency of predictions make them an attractive choice for regulatory decision support, particularly when used in combination with a more traditional 'read-across' approach to risk assessment. Furthermore, the scientific community has adopted a set of principles outlined by the Organization for Economic Cooperation and Development²⁷ for the validation of (Q)SARs that provide clearer expectations of what model characteristics are required for their use in a regulatory context. (Q)SAR tools can be very powerful when used appropriately and represent an increasingly reliable method for supporting the safety assessment of drugs and drug products.

Current Options for Carcinogenicity Assessment of Pharmaceuticals

ICH Guideline S1—Study options for small molecule pharmaceuticals. The current standard carcinogenicity study options for small-molecule pharmaceuticals are described in the 1997

ICH guideline S1B, "Testing for Carcinogenicity of Pharmaceuticals."¹⁸ Historically, the regulatory requirements for the assessment of carcinogenic potential of pharmaceuticals consisted of long-term studies in 2 rodent species, usually rats and mice. This approach involves relatively large number of animals to account for reduced survival and appropriate statistical evaluations and is an expensive endeavor. Over time, investigations demonstrated that it is possible to provoke a carcinogenic response in rodents by a diversity of experimental procedures that may or may not be relevant to human risk assessment. Therefore, the practice of requiring long-term carcinogenicity studies in 2 species was examined by an Expert Working Group of the ICH to determine whether it could be reduced without compromising human safety.

On the basis of that examination,⁷ the final ICH S1B recommended a strategy in 1997 comprised of a single long-term rodent carcinogenicity study, with the addition of one other supplementary study to provide additional information that is not readily available from the long-term assay. In the absence of clear evidence favoring one species, the guideline recommends the use of rats for the long-term assay. The additional study can be either a short- or medium-term *in vivo* rodent test system or a second long-term carcinogenicity study in another rodent species.

Use of transgenic mice. Section 4.2.2 of the ICH S1B guideline states that "additional *in vivo* tests for carcinogenicity" can be used as the second carcinogenicity assay and can include the use of transgenic mice. Studies in transgenic mice offer the advantage of requiring fewer animals per dose group (25 per sex per group) compared with traditional 2-y studies (60 animals per sex per group). This change in guidance was enabled by an effort coordinated by the International Life Sciences Institute and the Health and Environmental Science Institute in which 21 well-characterized materials, primarily pharmaceuticals, were tested in a number of transgenic mouse strains.²⁹ The results of these studies suggested that the transgenic models could correctly identify human carcinogens and noncarcinogens. However, a subsequent study²⁸ found that transgenic models miss some known and probable human carcinogens, which were only detected when the transgenic model was coupled with the 2-y rat study. The selection of the specific test method should be based on the overall weight of evidence (WOE) for the assessment of carcinogenic potential.

Despite the flexibility in approaches described under the current ICH S1B guidance, the majority of proposed carcinogenicity assessment programs received by the FDA prior to 2013 consisted of a 2-y study protocol in mice in addition to a 2-y study protocol in rats. Indeed, from 2002 to 2013, only about 25% of carcinogenicity assessments in mice were conducted in transgenic strains, although this percentage was approximately 35% during 2010 to 2013. In 2013, a slight majority of proposed mouse study protocols included an alternative mouse model. Since 2005, more than 80% of the transgenic study protocols received by the FDA were developed using the TgHras2 model (internal FDA database). Figure 1 summarizes the FDA experience with alternative carcinogenic model protocols during 2002 through 2013.

The uncertainty of how a regulatory agency would view a positive tumor response in a transgenic model, which is portrayed as a better predictor of human risk than are nontransgenic models,^{6,26} may be one explanation for the apparent resistance to the increased adoption of transgenic models by the pharmaceutical industry. In addition, the continued requirement of a 2-y study in rats may temper the potential reduction in animal

Year	Number alternative model protocols	p53 ±	Tg.AC	TgHras2	Other models
2002	10	5	4	1	0
2003	26	9	11	3	3
2004	17	7	3	6	1
2005	14	1	0	13	0
2006	13	5	1	6	1
2007	16	5	1	9	1
2008	16	3	1	12	0
2009	6	0	1	5	0
2010	14	2	0	12	0
2011	19	0	1	17	1
2012	12	0	0	9	3
2013	23	0	0	23	0

Figure 1. FDA experience with alternative mouse carcinogenicity models. The experience to date in 2014 is comparable to that in 2013.

numbers and savings in time, expense, and risk of using transgenic mice as the second species for evaluation.

ICH Guidelines S6 and S6(R1)—WOE approaches for biotechnology-derived products. ICH guidance S6, "Preclinical Safety Evaluation of Biotechnology-derived Pharmaceuticals," and its 2012 addendum, ICH S6(R1),^{17,19} recommend the use of a WOE approach for biotechnology-derived products when a carcinogenicity assessment is warranted. The guidance states that standard carcinogenicity bioassays generally are considered inappropriate for biotechnology-derived pharmaceuticals. However, product-specific assessments may be needed depending on the treatment duration, patient population, or biologic activity of the product. In specific cases, a single rodent species may be considered.

When an assessment of carcinogenic potential is warranted, a sponsor should design a strategy to address the potential hazard. The strategy could be based on a WOE approach, including a review of relevant data from sources such as published data, information on class effects, target biology and mechanism of action, *in vitro* data, and data from previously conducted chronic toxicity studies and clinical trials. In many cases, this type of assessment adequately addresses carcinogenic potential and precludes the need to conduct additional nonclinical studies. Rodent bioassays are not warranted when the WOE supports a concern regarding carcinogenic potential (for example, immunosuppressives and growth factors). Conversely, when the WOE assessment does not suggest carcinogenic potential, no additional nonclinical testing is recommended. In all, these recommended approaches reduce the unnecessary conduct of formal carcinogenicity studies when alternate strategies are available to address potential safety concerns for biotechnology-derived products.

A Path Forward on Carcinogenesis Testing of Pharmaceuticals

Familiarity with the conduct and limitations of the 2-y rodent bioassay, the predictable regulatory expectations across regions, and the acceptance of assay results by regulatory agencies helped to establish the 2-y rodent bioassay as the contemporary standard for assessing the carcinogenic potential of pharmaceuticals.

Alternative approaches often aim to eliminate the 2-y rat study altogether, with a notable exception¹⁵ to start dosing *in utero* in an attempt to encompass a greater portion of the rodent's lifespan exposed to drug. Most hypothesize that short-term toxicology findings can adequately predict a positive tumor outcome in 2-y rat studies, rendering the 2-y study unnecessary. Support for this idea is found in organ-specific assessments that demonstrate a positive correlation between short-term toxicologic changes to long-term tumor outcome for the liver, lung, and kidneys.^{5,8} This approach is potentially useful for predicting a tumor response in these specific organs; however, an analysis by the FDA found that findings from short-term studies (for example, 3-mo toxicology studies,

genotoxicity, metabolism, and pharmacology) poorly predict long-term tumor outcome on a whole-animal basis,²¹ which is the more relevant endpoint from the perspectives of risk assessment and public health. Other approaches, including epidemiologic methods, gene expression signatures, and signaling pathway analysis, leverage known modes of human carcinogenesis and may ultimately aid in predicting human risk rather than the tumor outcome in rodents.^{1,11,14}

An alternative approach to assessing carcinogenic potential that has gained particular attention was published by a consortium of interested pharmaceutical companies.³¹ Termed NegCarc (Negative for Endocrine, Genotoxicity, and Chronic Study Associated Histopathologic Risk Factors for Carcinogenicity), this proposal also is predicated on the hypothesis that outcomes from studies substantially less than a 2-y duration can adequately predict tumor outcome from long-term cancer bioassays. Whereas other efforts focused on *in vitro* or 3-mo short-term toxicologic findings indicative of a positive tumor outcome in rats,^{5,8} NegCarc proposed that the absence of a specific set of toxicologic findings adequately predicts the absence of drug-induced tumors in a 2-y rat bioassay. This conclusion was based on the analysis of 182 marketed and unmarketed pharmaceutical compounds voluntarily contributed by participating companies. Provided that a compound 1) did not result in "histopathologic risk factors of rat neoplasia" in a 6- or 12-mo toxicology study, 2) tested negative in all assays of the ICH S2 genotoxicity battery, and 3) did not perturb hormonal function by a WOE analysis, one could predict with 82% accuracy that such a compound would not produce tumors in a 2-y rat study.³¹ For the 18% of compounds where the prediction of a negative tumor outcome was wrong (that is, false negatives wherein tumors were observed in the 2-y rat study), none were interpreted as presenting a cancer risk to human subjects. If regulatory agencies agreed to waive the 2-y rat study for compounds with a NegCarc prediction of a negative tumor outcome, the authors estimated that approximately 40% of 2-y rat studies could be omitted from drug development programs without compromising patient safety.³¹ Recognizing the importance of this proposal and its potential effect on drug development, the FDA initiated a study with the intent of applying NegCarc as described in the Sistare publication³¹ to an independent set of compounds from FDA's archive. Analysis of 60 additional marketed and unmarketed compounds yielded a similar rates of negative predictivity and false negatives.¹⁶ Furthermore, the Japan Pharmaceutical Manufacturers Association reported similar results from analysis of 65 compounds, although the degree of compound overlap with the other datasets is uncertain.¹⁶ Nevertheless, 3 independently constructed datasets yielded consistent predictive properties of NegCarc, increasing confidence of a similar outcome should NegCarc be applied in the same manner to current and future drug development programs.

A primary obstacle to regulatory adoption of NegCarc arises from the definition of the 3 short-term toxicologic findings, or criteria, and how they are applied to determine the necessity of a 2-y rat study. Short-term toxicologic findings, as defined under NegCarc, that would otherwise be considered of minimal relevance to carcinogenic risk would nonetheless provide the rationale for conducting a 2-y rat study or denying a waiver request. For example, a drug-related positive result in a single assay of the genotoxicity battery or hepatocellular hypertrophy induced by a drug extensively metabolized by rat liver or a drug-related increase in thyroid-stimulating hormone would be individually sufficient to trigger the conduct of a 2-y rat study. Indeed, among the cases where NegCarc indicated the need

for a 2-y study, approximately half were initiated due to a positive finding in only 1 of the 3 defined criteria.³¹ This empirical approach is necessary to achieve the predictive properties of NegCarc; if the criteria are narrowed by excluding toxicologic changes perceived to be irrelevant to a tumor risk, its predictive properties necessarily change in an undefined manner. In addition, regulatory agencies may frequently be placed in the untenable situation of denying a 2-y rat study waiver request for reasons that are scientifically unjustified.

An equally concerning obstacle to adopting NegCarc,³¹ as proposed, is the presumption that false-negative cases, or predicting as negative 'true' rodent carcinogens, would not have implications for patient safety. In FDA's experience, when an unexpected cancer signal is uncovered in the course of phase 3 clinical trials or in postmarket experience with an approved drug, the results of 2-y rodent studies are among the first data reexamined in assessing plausibility of the clinical finding. This observation alone demonstrates that sponsors and regulatory agencies do find value from 2-y rat studies in specific situations. In the absence of such information in cases where a waiver was granted based on a prediction of minimal cancer risk, a false negative would be suspected, and a 2-y study potentially would be mandated as part of the strategy of investigating the clinical signal. In this context, posthoc discovery of a corroborative carcinogenic signal in rodents presents challenges in assessing the appropriate action to be taken by the sponsor and the regulatory agency.

Empirically derived decisions based on NegCarc data frequently could arise from chance associations between the toxicologic trigger and the observed tumor outcome in the 2-y rat study, yet this cost is necessary to achieve reasonable negative predictivity with this method. Decades of experience with 2-y rodent bioassays and associated studies have shown that some classes of nongenotoxic compounds result in the same tumor profile with reasonable reproducibility. For example, β 2-adrenergic receptor agonists often result in mesovarian leiomyomas, drugs with dopamine antagonist activity result in mammary neoplasia, and agonists of peroxisome proliferator-activated receptor isotypes frequently result in some combination of bladder, adipose, blood vessel, and liver neoplasms in rodents.^{10,30} Rational explanations for many drug class-related tumor responses in rodents are often found from an understanding of the pharmacologic and pharmacokinetic properties of the drug class, as supported by mechanistic studies. This information was absent in the NegCarc datasets and therefore analysis of its potential use in predicting tumor responses was not possible. With the cooperation of Sistare and colleagues,³¹ the European Medicines Agency, and the Pharmaceutical and Medical Devices Agency of Japan, the FDA unblinded the NegCarc dataset regarding compound identification and drug class. Upon inspection, it was determined that a drug's pharmacologic properties could reasonably account for, and thus predict, the rodent tumor response in many cases where NegCarc failed (that is, false negatives), a determination that has implications not only for better detection of false negatives but potentially for all compound classifications in the dataset.

This insight led to additional discussions between drug regulatory agencies (DRA) and pharmaceutical associations as part of an Expert Working Group of the International Conference on Harmonization. The working hypothesis, summarized in a Regulatory Notice Document¹⁶ is that more rational predictions of carcinogenic potential for rodents and potentially for human subjects could be achieved by considering pharmacologic

properties of a given investigational drug in addition to the toxicologic endpoints described in NegCarc.¹⁶ Furthermore, it has been suggested that *in silico* models could provide additional supporting information to this end.²² But unlike their empirical application in the Sistare publication,³¹ the contribution of the short-term toxicologic endpoints would be interpreted in the context of the drug's pharmacology and pharmacokinetics in assessing whether a reasonable prediction could be made regarding carcinogenic risk. The Expert Working Group advanced the following construct in the notice¹⁶ that attempts to describe possible outcomes from a WOE evaluation for carcinogenic risk, defining the circumstances under which a 2-y rat study is needed to support drug development programs. Assuming sufficient evidence, it may be possible to waive the need for a 2-y rat study when a compound fits the following criteria:

Category 1—a product is likely to be tumorigenic in humans;

Category 3a—a product is likely to be tumorigenic in rats but not in humans through prior established and well-recognized mechanisms known to be human irrelevant; or

Category 3b—a product is not likely to be tumorigenic in both rats or humans.

When sufficient evidence is lacking or describes an equivocal picture, then a product would be considered Category 2, wherein a 2-y rat study is needed and could add value to assessing carcinogenic potential. In cases where the WOE evaluation is supportive of omitting the 2-y rat study, the Expert Working Group proposed that either a 2-y mouse or a transgenic mouse carcinogenicity study would still be needed in most cases.¹⁶

The complexities in assessing carcinogenic risk of pharmaceuticals precludes, to some degree, simply prescribing what constitutes sufficient evidence in support of these categorizations and will need to be informed by experience. To this end, it is instructive to recognize that ICH S6(R1) does not overtly prescribe specific elements to include in a WOE assessment for carcinogenic potential of biotechnology-derived pharmaceuticals. Yet regulatory agencies are able to make decisions regarding the need for 2-y rodent studies based on these WOE arguments, even for biotechnology-derived pharmaceuticals where a 2-y rodent study is feasible. But it is recognized that small-molecule drugs will present a greater challenge to any WOE assessment of carcinogenic risk compared with biotechnology-derived products, because these two types of product differ in some important aspects, particularly target specificity and metabolism. Of some advantage, current *in silico* models for carcinogenicity are designed for use on small-molecule compounds, whereas they are not predictive for biotechnology-derived pharmaceuticals. The added complexity of small-molecule drugs suggests that a WOE approach might be most successful in identifying only those drugs with the clearest evidence of a human carcinogenic hazard (Category 1) and those that have minimal or no human risk (Categories 3a and 3b). It is reasonable to expect that the level of evidence supporting the latter would be higher than for the former.

Identifying cases where evidence is insufficient or equivocal will predictably be an easier task, thus supporting Category 2 status and the conduct of a 2-y rat study. Indeed, the majority of compounds likely fall in this category. But here, too, a WOE approach early in the drug development program, possibly including *in silico* assessment, could afford greater flexibility in identifying a more rational and informative path forward for addressing the carcinogenic potential of a drug, depending on the issues specific to that drug.

Moving toward a WOE assessment of carcinogenic risk for small-molecule products is not a novel concept^{7,24} but presents

formidable challenges. The suggestion that carcinogenic potential could be assessed adequately for some pharmaceuticals without conducting a 2-y rat study comes from retrospective analysis of the NegCarc dataset and others.^{9,16,34} However, there is no evidence that prospectively applying a set of WOE criteria will acceptably predict the outcome and value of a 2-y rat study or improve assessment of carcinogenic risk to human subjects. In addition, the assessment of carcinogenic potential is harmonized currently by the ICH S1 guidelines, which have successfully set expectations for the pharmaceutical industry and DRA worldwide. Moving toward an approach where DRA draw decisions from review of a sponsor's WOE assessment will likely result in some degree of disharmony, because the DRA may come to different conclusions regarding the necessity of a 2-y rat study for the same investigational pharmaceutical.

Addressing these and other challenges associated with a WOE approach is possible. The ICH Expert Working Group recently has described a prospective study whereby WOE assessments, or carcinogenicity assessment documents, are submitted voluntarily from pharmaceutical companies to DRA prior to initiating 2-y rat studies or before the results of an ongoing 2-y rat study are known.¹⁶ These documents address the toxicologic and pharmacologic aspects of the drug and any other information the contributing sponsor considers pertinent to prospectively categorizing their compound as Category 1, 2, or 3a/3b. The DRA then independently evaluate the carcinogenicity assessment documents in a blinded manner to arrive at their own categorization for the compound, and record whether they agree or disagree with the sponsor's chosen category. The DRA are blinded to the identity of the compound, the contributing sponsor, and to each other's deliberations prior to deciding on the categorization of the compound. Predictions then would be checked against the actual tumor outcome and value of the 2-y rat study, once completed.

The experience gained and data generated from this prospective study could best address some critical issues that prior retrospective studies could not address as well as provide a more objective assessment of the proposed approach. The results would demonstrate just how accurately one can predict the tumor outcome of a 2-y rat study based on a WOE assessment. And, perhaps more importantly, those elements that were most useful—or most misleading—in making those predictions would be identified and the information used to improve future predictive assessments. Differences in interpretation of data and their potential consequences would be highlighted by determining the degree of concordance between the DRA and sponsors regarding the categorical prediction in the carcinogenicity assessment documents. The degree of discordance between the DRA regarding the necessity for a 2-y rat study for the same pharmaceutical would illuminate the degree of disharmony introduced into drug development across regulatory regions with a WOE approach. Although some discordance between DRA and sponsors is expected, as is some discordance between DRA on the necessity of a 2-y rat study, a high degree of discordance in either measure would require remedy prior to the adoption of any WOE-based approach.

Ideally, results from this prospective study would help define the circumstances under which a 2-y rat study meaningfully contributes to understanding the carcinogenic potential of a compound. Equally important, the circumstances under which a 2-y rat study did not prove useful in understanding a compound's carcinogenic potential would be captured as well. Recognizing such cases would greatly aid in moving assessment of carcinogenic potential of pharmaceuticals away from a

screening activity and toward a more rational and informative assessment framework.

Conclusions and the Future of Carcinogenicity Testing

Efforts to develop strategies to better predict human cancer risk associated with pharmaceuticals and reduce the use of animals overall are progressing. This advance is evident when one considers the strategies described in the ICH S1 guidelines and the current efforts in building improved in silico methods and refining in vivo strategies. Flexibility exists in the contemporary approaches to assessing carcinogenicity of pharmaceuticals: WOE evaluations are encouraged for biologic products even when 2-y studies are feasible and shorter-term transgenic mouse studies and other alternatives described in S1B may be proposed in place of a 2-y mouse study for small molecules. In some circumstances, evaluation in one species can suffice and the timing of the study delayed to the postapproval period. Looking ahead, current deliberation at the ICH could further extend flexibility by recommending a WOE analysis in place of the 2-y rat study under certain conditions, as a complement to a transgenic mouse or 2-y mouse study.

A key shortcoming of many efforts to date is the use of data from 2-y bioassays as the 'gold standard' despite their recognized limitations. Many positive responses in rodents are unlikely to represent human risk, but the reverse is true as well: a negative response in rodents may not always translate to a lack for human risk. Some drugs can only be tested in rodents at a fraction of human clinical exposures, and the negative results in those studies could represent false negatives. For the purposes of in silico modeling, standardized data sets for human carcinogenicity are too sparsely populated to develop robust (Q) SAR models; however, newer rodent carcinogenicity datasets capturing specific tumor type information may provide a path forward to elucidating the difference between human-relevant and nonrelevant predictions of rodent carcinogenicity.

Perhaps the best way to make progress in identifying human cancer risk from pharmaceutical exposures is to study humans. Phase 1 clinical trials often involve tolerability studies in which participants are dosed at levels well beyond what will become standard clinical dosages. Late-stage studies can involve large numbers of patients exposed for weeks, months, and sometime years. Study participants can be evaluated prior to drug dosing and at intervals during the course of the trials. Obviously, the tissues that can be evaluated are limited but may still provide a window into biologic changes that presage increased risk for cancer. Accepted and validated endpoints necessary to monitor potential cancer risk in clinical trial subjects remain exploratory and are unavailable; therefore animal studies will necessarily remain the standard by which cancer risks for new pharmaceuticals is assessed. Yet efforts to develop robust methods for clinical monitoring of potential cancer risks from pharmaceuticals may be more fruitful than building a better rodent bioassay.

References

1. Adami HO, Berry CL, Breckenridge CB, Smith LL, Swenberg JA, Trichopoulos D, Weiss NS, Pastoor TP. 2011. Toxicology and epidemiology: improving the science with a framework for combining toxicological and epidemiological evidence to establish causal inference. *Toxicol Sci* 122:223–234.
2. Ames BN, McCann J, Yamasaki E. 1975. Methods for detecting carcinogens and mutagens with the *Salmonella*-mammalian-microsome mutagenicity test. *Mutat Res* 31:347–364.

3. Benigni R. 2005. Structure–activity relationship studies of chemical mutagens and carcinogens: mechanistic investigations and prediction approaches. *Chem Rev* 105:1767–1800.
4. Benigni R, Giuliani A, Franke R, Gruska A. 2000. Quantitative structure–activity relationships of mutagenic and carcinogenic aromatic amines. *Chem Rev* 100:3697–3714.
5. Boobis AR, Cohen SM, Doerrer NG, Galloway SM, Haley PJ, Hard GC, Hess FG, Macdonald JS, Thibault S, Wolf DC, Wright J. 2009. A data-based assessment of alternative strategies for identification of potential human cancer hazards. *Toxicol Pathol* 37:714–732.
6. Boverhof DR, Chamberlain MP, Elcombe CR, Gonzalez FJ, Heflich RH, Hernandez LG, Jacobs AC, Jacobson-Kram D, Luijten M, Maggi A, Manjanatha MG, Benthem JV, Gollapudi BB. 2011. Transgenic animal models in toxicology: historical perspectives and future outlook. *Toxicol Sci* 121:207–233.
7. Cohen SM. 2001. Alternative models for carcinogenicity testing: weight of evidence evaluations across models. *Toxicol Pathol* 29 Suppl:183–190.
8. Cohen SM. 2010. Evaluation of possible carcinogenic risk to humans based on liver tumors in rodent assays: the 2-year bioassay is no longer necessary. *Toxicol Pathol* 38:487–501.
9. Contrera JF, Jacobs AC, DeGeorge JJ. 1997. Carcinogenicity testing and the evaluation of regulatory requirements for pharmaceuticals. *Regul Toxicol Pharmacol* 25:130–145.
10. El Hage J. [Internet]. 2005. US Food and Drug Administration, Endocrinologic and Metabolic Drugs Advisory Committee NDA 21-865. Pharmacology–toxicology review. [Cited 6 Feb 2014]. Available at: www.fda.gov/ohrms/dockets/ac/05/slides/2005-4169S2_00-Slide-Index.htm
11. Fielden MR, Adai A, Dunn RT, Predictive Safety Testing Consortium, Carcinogenicity Working Group. 2008. Interlaboratory evaluation of genomic signatures for predicting carcinogenicity in the rat. *Toxicol Sci* 103:28–34.
12. Franke R, Gruska A, Giuliani A, Benigni R. 2001. Prediction of rodent carcinogenicity of aromatic amines: a quantitative structure–activity relationships model. *Carcinogenesis* 22:1561–1571.
13. Franke R, Gruska A, Bossa C, Benigni R. 2010. QSARs of aromatic amines: identification of potent carcinogens. *Mutat Res* 691:27–40.
14. Guyton KZ, Kyle AD, Aubrecht J, Cogliano VJ, Eastmond DA, Jackson M, Keshava N, Sandy MS, Sonawane B, Zhang L, Waters MD, Smith MT. 2009. Improving prediction of chemical carcinogenicity by considering multiple mechanisms and applying toxicogenomic approaches. *Mutat Res* 681:230–240.
15. Huff J, Jacobson MF, Davis DL. 2008. The limits of 2-year bioassay exposure regimens for identifying chemical carcinogens. *Environ Health Perspect* 116:1439–1442.
16. International Conference on Harmonization. [Internet]. 2013. Regulatory Notice Document: proposed change to rodent carcinogenicity testing of pharmaceuticals. [Cited 6 Feb 2014]. Available at: www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Safety/S1/S1_Regulatory_Notice_Document_8.Aug.2013.pdf
17. International Conference on Harmonization. [Internet]. 1997. S6: preclinical safety evaluation of biotechnology-derived pharmaceuticals. [Cited 6 Feb 2014]. Available at: <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM074957.pdf>
18. International Conference on Harmonization. [Internet]. 1998. S1B: testing for carcinogenicity of pharmaceuticals. [Cited 6 Feb 2014]. Available at: <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM074916.pdf>
19. International Conference on Harmonization. [Internet]. 2012. S6: addendum to preclinical safety evaluation of biotechnology-derived pharmaceuticals. [Cited 6 Feb 2014]. Available at: <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM074957.pdf>
20. International Conference on Harmonization. [Internet]. 2013. M7: assessment and control of DNA reactive (mutagenic) impurities in pharmaceuticals to limit potential carcinogenic risk. [Cited 6 Feb 2014]. Available at: <http://www.ich.org/fileadmin/>

- Public_Web_Site/ICH_Products/Guidelines/Multidisciplinary/M7/M7_Step_2.pdf
21. **Jacobs A.** 2005. Prediction of 2-year carcinogenicity study results for pharmaceutical products: how are we doing? *Toxicol Sci* **88**:18–23.
 22. **Kruhlak NL, Benz RD, Zhou H, Colatsky TJ.** 2012. (Q)SAR modeling and safety assessment in regulatory review. *Clin Pharmacol Ther* **91**:529–534.
 23. **LeBoeuf RA, Kerckaert GA, Aardema MJ, Gibson DP, Brauning R, Isfort RJ.** 1996. The pH 6.7 Syrian hamster embryo cell transformation assay for assessing the carcinogenic potential of chemicals. *Mutat Res* **356**:85–127.
 24. **MacDonald JS.** 2004. Human carcinogenic risk evaluation, part IV: assessment of human risk of cancer from chemical exposure using a global weight-of-evidence approach. *Toxicol Sci* **82**:3–8.
 25. **Matthews EJ, Contrera JF.** 1998. A new highly specific method for predicting the carcinogenic potential of pharmaceuticals in rodents using enhanced MCASE QSAR–ES software. *Regul Toxicol Pharmacol* **28**:242–264.
 26. **Morton D, Sistare FD, Nambiar PR, Turner OC, Radi Z, Bower N.** 2013. Regulatory forum commentary: alternative mouse models for future cancer risk assessment. *Toxicol Pathol* **42**:799–806 .
 27. **Organization for Economic Co-operation and Development.** 2007. Guidance document on the validation of (quantitative) structure–activity relationship [(Q)SAR] models. <http://www.oecd.org/env/ehs/risk-assessment/guidancedocumentsandreportsrelatedtoqsars.htm>
 28. **Pritchard JB, French JE, Davis BJ, Haseman JK.** 2003. The role of transgenic mouse models in carcinogen identification. *Environ Health Perspect* **111**:444–454.
 29. **Robinson DE, MacDonald JS.** 2001. Background and framework for ILSI’s collaborative evaluation program on alternative models for carcinogenicity assessment. *Toxicol Pathol* **29 Suppl**:13–19.
 30. **Silva Lima B, Van der Laan J.** 2000. Mechanisms of nongenotoxic carcinogenesis and assessment of the human hazard. *Regul Toxicol Pharmacol* **32**:135–143.
 31. **Sistare FD, Morton D, Alden C, Christensen J, Keller D, Jonghe SD, Storer RD, Reddy MV, Kraynak A, Trela B, Bienvenu JG, Bjurström S, Bosmans V, Brewster D, Colman K, Dominick M, Evans J, Hailey JR, Kinter L, Liu M, Mahrt C, Marien D, Myer J, Perry R, Potenta D, Roth A, Sherratt P, Singer T, Slim R, Soper K, Fransson-Steen R, Stoltz J, Turner O, Turnquist S, van Heerden M, Woicke J, DeGeorge JJ.** 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.
 32. **Stavitskaya L, Kruhlak NL, Cross KP, Minnier BL, Bower DA, Chakravarti S, Saiakhov RD, Benz RD.** 2012. Development of improved in silico models for predicting rodent carcinogenicity. American College of Toxicology 33rd Annual Meeting Program Book, Abstract P217, p 94 Orlando, FL.
 33. **Tennant RW, Margolin BH, Shelby MD, Zeiger E, Haseman JK, Spalding J, Caspary W, Resnick M, Stasiewicz S, Anderson B, Minor R.** 1987. Prediction of chemical carcinogenicity in rodents from in vitro genetic toxicity assays. *Science* **236**:933–941.
 34. **Van Oosterhout JPJ, Van der Laan JW, De Waal EJ, Olejniczak K, Hilgenfeld M, Schmidt V, Bass R.** 1997. The utility of 2 rodent species in carcinogenic risk assessment of pharmaceuticals in Europe. *Regul Toxicol Pharmacol* **25**:6–17.