frontiers in
PHARMACOLOGY

In vitro toxicity testing in the twenty-first century

*Erwin L. Roggen**

Novozymes AS, Bagsvaerd, Denmark

The National Research Council (NRC) article "*Toxicity Testing in the 21st Century: A vision and A Strategy*" (National Research Council, 2007) was written to bring attention to the application of scientific advances for use in toxicity tests so that chemicals can be tested in a more time and cost efficient manner while providing a more relevant and mechanistic insight into the toxic potential of a compound. Development of tools for *in vitro* toxicity testing constitutes an important activity of this vision and contributes to the provision of test systems as well as data that are essential for the development of computer modeling tools for, e.g., system biology, physiologically based modeling. This article intends to highlight some of the issues that have to be addressed in order to make *in vitro* toxicity testing a reality in the twenty-first century.

Keywords: *in vitro* toxicology, human-specific methods, pathways of toxicity, markers of toxicity, adverse responses

HUMAN-SPECIFIC METHODS – THE CHALLENGES

In vitro toxicity testing should build upon test models that are relevant for the species to be protected. Proper test development requires well defined test compounds with high quality *in vivo* data (gold standard) and cell systems that mimic *in vitro* the key events that are known to occur *in vivo*.

Outside the pharmaceutical industry, adequate gold standards based upon human data are very rare. Consequently, human cell-based tests are often developed against gold standards of animal origin and may not reflect events occurring in humans after exposure to a toxic compound. One well established example is nickel-induced contact dermatitis. When tested on mice, there is no evidence for nickel inducing contact hypersensitivity. However, there is ample evidence provided by both *in vitro* tests based upon human cells as well as human clinical data demonstrating that nickel induces contact dermatitis (Schmidt et al., 2010). One way to overcome this hurtle is to acquire a solid and in-depth understanding of the mode of action and mechanisms of action driving a toxic response in humans. Such an understanding may provide the confidence that is required to make the leap from animal experiments to *in vitro* human cell-based toxicology for protecting humans. Referring to the example of nickel, in-depth mechanistic studies have demonstrated that species-specific differences in the response to nickel are related to differences between the amino acid sequences of mouse and human Toll-like receptor 4 (Schmidt et al., 2010).

In general, the majority of the currently available cell-based models suffer from a series of limitations (e.g., reduced metabolic competence, cancer cells) which future research and development need to address (Prieto et al., 2006; Hartung, 2007). The lack of a regular supply of human tissue jeopardizes the availability of a number of cell-based tests (e.g., liver, lung, brain). An obvious solution is the use of sustainable human cell lines or human stem cell technology. However, both cell types face the issue of *in vivo*-like functionality (or lack thereof). For those tissues where availability is a less of a problem (e.g., skin), primary cells are used. An important limitation of primary cells is donor-to-donor variability, which in many cases affects the reproducibility of the test in question. Also here, sustainable cell lines or stem cells are a solution.

Whether *in vitro* tests are based on primary cells, immortalized (e.g., SV40 transformation) and cancer-derived cell lines, stem cells, or reconstituted tissue cultures, it is important to have *in vitro* systems that adequately mimic key events of the *in vivo* mechanisms of action triggered in humans upon exposure to a toxic compound. Indeed, cells or tissue may no longer exhibit relevant *in vivo*-like functionality with respect to the endpoint and type of compounds to be tested. In other words, they may no longer express mechanisms that *in vivo* are required for a compound to be toxic (giving false negative responses), or they may express mechanisms and are not active *in vivo* (resulting in false positive responses) in a healthy individual. Thus, proper functional characterization of cells, cell lines (including stem cells), and tissue is required before initiation of test development. There is no need to say that defining whether or not a cell-based test reveals adequate *in vivo* functionality is a difficult task requiring solid understanding of the physiological mechanisms occurring in humans *in vivo* and in cells *in vitro*. It has to be stressed also that the definition of *in vivo*-like functionality is dependent on the question (e.g., impact of a compound on an endpoint such as cytokine release or key event such as a pathway) that the cell assay has to answer.

Since cell functionality *in vivo* is driven by the microenvironment surrounding the cell, as well as by cell–cell interactions, development of *in vivo*-like *in vitro* tests requires urgently a better understanding of the impact of the microenvironment and cell–cell interactions on the mechanisms driving cell differentiation, dedifferentiation, and responsiveness to, e.g., xenobiotics. This understanding is required to boost the development of well-characterized human-derived proteins (e.g., for establishing defined culture media), cell lines and cells (including stem cells), organ cultures, and tissues for *in vitro* modeling of *in vivo*-relevant toxicological events.

FROM IDENTIFICATION OF PATHWAYS TO KEY EVENTS AND MARKERS OF TOXICITY

In vitro toxicity testing should build upon an in-depth understanding of the physiological processes related to toxicological endpoints, and to find the key pathways and components of these

pathways involved in the responses to toxin exposure. Based upon the experiences within the field of carcinogenicity, it is expected that the number of relevant pathways is limited to tens to hundreds (Johnson et al., 2004; Van Delft et al., 2004). Thus, high throughput and content screening tests are needed using human-specific assays. In this context, the models may not necessarily be derived from the target organ, rather should simply demonstrate the presence of the pathway or mechanism of interest and the effect of a chemical upon it.

It may be possible to acquire further insight into human *in vivo* mechanisms and pathways, and to assess the relevance of the *in vitro* identified mechanisms, by implementing tools used by the pharmaceutical industry in human clinical studies (e.g., micro-dosing and tracing studies). A better understanding of the mechanisms and pathways involved may in the end allow for data extrapolation from a healthy to a diseased state.

There is a general need for markers and marker profiles with adequate power to predict toxicity (including potency) and, in the case of pharmaceuticals, efficacy. For several diseases (e.g., allergy, chronic diseases, cancer), specific clusters of genes have been identified and evaluated. Gene-cluster modeling has increased our understanding of the mechanisms of action driving the clinical conditions, and diagnostic markers have been identified (Gohlke et al., 2009). The relevance of these markers for toxicity testing remains to be established. Gene-cluster modeling before and after exposure of human-specific *in vitro* test systems to toxic and non-toxic compounds has been performed in an effort to identify new markers and marker profiles for toxicity. Progress has been made, but the resulting markers and marker signatures remain to be optimized and adapted for prediction of a specific endpoint related to a specific clinical condition.

DEFINING ADVERSITY

To date, human risk assessment is based upon thresholds defining "no effect levels" (chemical and cosmetic industry) or "no adverse effect levels" (pharmaceutical industry) in animal studies. Defining a threshold for humans based upon data from animal experiments has been and still is challenging, and often leads to false positive results. From an industrial point of view, a high rate of false positives has to be avoided as this leads to the elimination of often promising compounds. From the risk assessment point of view, false negative results threaten human safety and should be avoided.

The current animal-based perception of "no adverse effect levels" has been challenged by the high sensitivity of the emerging techniques (e.g., genomics, proteomics, metabonomics) making it possible to detect responses at very low doses of a compound. The consequences are very evident in the area of genotoxicity, where thus far any effect of a non-pharmaceutical compound on biological *in vivo* systems results in a "no go." Indeed, the high sensitivity has made it possible to detect changes induced by other factors than chemical exposure (e.g., changes in nutrient concentrations and pH, cell cycle, and aging). In addition, exposure to low dose of a chemical will induce detectable changes not leading to the demise of the cell *per se* but causing changes in, e.g., signal pathways to counteract the effect of the chemical. These events should not be equated to a high dose effect which causes irreversible cell injury.

Thus, the new technologies have made it imperative to understand the mechanistic differences in adverse and adaptive responses to compound exposure.

Special attention should be given to chemicals to which humans are chronically exposed. Indeed, there is a growing concern about the impact of doses that cause adaptive responses when these doses are imposed on the system for longer periods.

Finally, responses may be modified by adjacent cells and tissues. The time line of exposure and responses may also differ. Tools that make it possible to address these issues (e.g., inter-connected cell culture systems, imaging techniques, interactomics, physiologically based pharmacokinetics) have to be implemented.

TOOLS FOR INCREASING PREDICTIVITY

To decrease the number of animals used for *in vivo* toxicity testing, the use toxicogenomics for identifying and/or dissecting the mechanisms of action of a test compound has been recommended.

Toxicogenomics can provides a library of generic expression profiles for different classes of toxicity that allows the characterization of an unknown compound based upon the profiles with which it fits. While genomics is used on a large scale for pathway analysis and marker identification, this concept has not yet been fully implemented in toxicity testing strategies and risk assessment.

Carcinogenicity testing is in this respect an interesting case study. The use of toxicogenomics for identifying the mechanisms of action of genotoxic and non-genotoxic carcinogens has been increasing over the past few years and there are now training sets for carcinogens and non-hepatotoxic non-carcinogens. The learnings of this case study should be also implemented on other toxicological endpoints (Johnson et al., 2004; Van Delft et al., 2004).

It can be anticipated that the integration of genomics, proteomics, and metabonomics data obtained from exposed and unexposed cellular or animal models, and clinical samples, will improve our understanding of the mechanisms of action of a test compound significantly (Hanahan and Weinberg, 2000). Furthermore, these data will help to establish relevant associations using newly developed computational technologies (e.g., systems biology).

INTEGRATED TESTING STRATEGIES

It is anticipated that a more in-depth understanding of the relation between toxicity and biological pathways will make it possible to prevent animal testing by using a combination of tests that individually represent key events of the mechanisms of action of toxicity and that allow for assessment of the potential of a test compound to affect these key events. *In vitro* and *in silico* methods can be used to accomplish this. If sufficient scientific justification is provided it may be possible to waive an animal test.

When selecting the battery of *in vitro* and *in silico* methods addressing key steps in the relevant biological pathways, it is important to employ standardized and internationally accepted tests. Each block should be producing data that are reliable, robust, and relevant (the alternative 3R elements) for assessing the specific

aspect (e.g., biological pathway) it is supposed to address. If they comply with these elements they can be used in integrated testing strategies.

To date there are no existing procedures and guidelines for putting together and validating such strategies. Obviously, this constitutes a hurdle for the implementation by regulatory agencies (Kinsner-Ovaskainen et al., 2009).

VALIDATION, IMPLEMENTATION, AND ACCEPTANCE

It is important to keep in mind that *in vitro* tests do not have fewer limitations than *in vivo* tests. Therefore, proof is needed that a new method is equal to or better than an existing *in vivo* traditional model. An added challenge is that since science is moving very quickly it is difficult to decide when a test is good enough to be a final test for risk assessment.

There is a need to incorporate new thinking into risk assessment. Regulators are receptive to new technologies but concrete data (e.g., mechanistic understanding and relevance) are needed to support their use. Data documentation should be comprehensive, traceable, and make it possible for other investigators to retrieve information as well as reliably repeat the studies in question regardless of whether the original work was performed to Good Laboratory Praxis (GLP) standards.

It is important to address in a systematic way the factors that are critical for assay reproducibility and reliability. An issue often faced while performing cell-based tests is intra- and inter-laboratory variability in spite of rigorous compliance with the Standard Operation Procedure (SOP). The reasons for this variability are often undefined but it is generally accepted that the causes include the cell cultures, analytical processing, technical error, and differences in qualitative judgment. Therefore, these parameters should be carefully addressed and standardized. Retrospective weight of evidence would be one tool for harmonizing how people perform specific tests and to assure good quality of the data. This would help to identify flaws in the analytical processes, technical error, and qualitative judgment. The exploitation by the *in vitro* testing community of emerging nano-biotechnologies facilitating the real time monitoring of cellular activity and processes reflecting the quality of the cell culture would provide objective tools for eliminating variations in the performance of cell-based tests.

REFERENCES

Gohlke, J. M., Thomas, R., Zhang, Y., Rosenstein, M. C., Davis, P. C., Murphy, C., Becker, K. G., Mattingly, C. J., and Portier, C. J. (2009). Genetic and environmental pathways to complex diseases. *BMC Syst. Biol.* 3, 46. doi: 10.1186/1752-0509-3-46

Hanahan, D., and Weinberg, R. A. (2000). The hallmarks of cancer. *Cell* 100, 57–70.

Hartung, T. (2007). Food for thought… on cell culture. *ALTEX Altern. Anim. Exp.* 24, 143–147.

Johnson, C. D., Balagurunathan, Y., Tadesse, M. G., Falahatpisheh, M. H., Brun, M., Walker, M. K., Dougherty, E. R., and Ramos, K. S. (2004). Unravelling gene–gene interactions regulated by ligands of the aryl hydrocarbon receptor. *Environ. Health Perspect.* 112, 403–412.

Kinsner-Ovaskainen, A., Akkan, Z., Casati, S., Coecke, S., Corvi, R., Dal Negro, G.,

De Bruijn, J., De Silva, O., Gribaldo, L., Griesinger, C., Jaworska, J., Kreysa, J., Maxwell, G., McNamee, P., Price, A., Prieto, P., Schubert, R., Tosti, L., Worth, A., and Zuang, V. (2009). Overcoming barriers to validation of non-animal partial replacement methods/integrated testing strategies: report of the EPAA-ECVAM workshop. *Altern. Lab. Anim.* 37, 437–444.

National Research Council. (2007). *Committee on Toxicity Testing and Assessment of Environmental Agents, Board on Environmental Studies and Toxicology, Institute for Laboratory Animal Research. Toxicity Testing in the 21st Century. A Vision and a Strategy.* Washington, DC: National Academies Press.

Prieto, P., Baird, A. W., Blaauboer, B. J., Ripoll, J. V. C., Corvi, R., Dekant, W., Dietl, P., Gennari, A., Gribaldo, L., Griffin, J. L., Hartung, T., Heindel, J. J.,

Hoet, P., Jennings, P., Marocchio, L., Noraberg, J., Pazos, P., Westmoreland, C., Wolf, A., Wright, J., and Pfaller, W. (2006). The assessment of repeated dose toxicity in vitro: a proposed approach. *Altern. Lab. Anim.* 34, 315–341.

Schmidt, M., Raghavan, B., Müller, V., Vogl, T., Fejer, G., Tchaptchet, S., Keck, S., Kalis, C., Nielsen, P., Galanos, C., Roth, J., Skerra, A., Martin, S. F., Freudenberg, M. A., and Goebeler, M. (2010). A crucial role for human Toll-like receptor 4 in the development of contact allergy to nickel. *Nat. Immunol.* 11, 814–819.

Van Delft, J. H. M., van Agen, E., van Breda, S. G. J., Herwijnen, M. H., Staal, Y. C. M., and Kleinjans, J. C. S. (2004). Discrimination of genotoxic from non-genotoxic carcinogens by gene expression profiling. *Carcinogenesis* 25, 1265–1276.