First Alternative Method Validated by a Retrospective Weight-of-Evidence Approach to Replace the Draize Eye Test for the Identification of Non-Irritant Substances for a Defined Applicability Domain

Thomas Hartung[1,2], Leon Bruner[3], Rodger Curren[4], Chantra Eskes[5], Alan Goldberg[1], Pauline McNamee[6], Laurie Scott[7] and Valérie Zuang[8]

[1]Johns Hopkins University, Bloomberg School of Public Health, Center for Alternatives to Animal Testing (CAAT), Baltimore, USA; [2]University of Konstanz, CAAT Europe, Konstanz, Germany; [3]The Procter & Gamble Company Needham Technical Center, Needham, USA; [4]Institute for In Vitro Sciences, Inc., Gaithersburg, USA; [5]Independent Consultant, Ispra, Italy; [6]The Procter & Gamble Company, London Innovation Centre, Egham, UK; [7]The Procter & Gamble Company, Cincinnati, USA; [8]In Vitro Methods Unit/ECVAM, Institute for Health and Consumer Protection, Joint Research Centre, European Commission, Ispra, Italy

Summary

A replacement alternative to the rabbit eye irritation test has been sought for many years. First published in 1944 by FDA toxicologist J. H. Draize, the test, now known as the Draize Eye Test, has been used extensively to assess eye safety. It has also been a focal point for concern regarding its animal use. In 1992, Molecular Devices developed the Cytosensor Microphysiometer (CM) technology, an automated potentiometric online measurement of pH changes in cells, and evaluated it also for chemically induced irritation.

The method was included in some of the six major validation studies for eye irritation from 1991-1997. The results for CM were inconclusive as were those from other tests evaluated as stand-alone methods to fully replace the animal test. In 2002, the European Centre for the Validation of Alternative Methods (ECVAM) started applying concepts from evidence-based medicine, and opened validation to retrospective meta-analysis. This activity was done in collaboration with US counterpart ICCVAM/NICEATM, and the European Cosmetics Association, Colipa.

After a new, comprehensive evaluation of the prior available data, the ECVAM scientific advisory committee (ESAC) has recently accepted the CM as capable of identifying non-irritants for testing limited to water-soluble surfactants and water-soluble surfactant-containing mixtures. This 25-year development is remarkable and instructive in many respects. The authors see this as opening the door, at last, for an end to the use of animals as a standard requirement for eye irritation. Here, several of the people critically involved in this processes have summarized the important aspects of this history.

Keywords: Draize rabbit eye test, alternative methods, validation, history, Cytosensor Microphysiometer

1 Brief introduction to eye irritation

The need for ocular irritancy testing became clear in the 1930s, when an untested eyelash product containing p-phenylene diamine was marketed in the US. Use of this and similar products led to sensitization of the external ocular structures, corneal ulceration, and vision loss (Wilhelmus, 2001). This in turn led to the passage of the United States Federal Food, Drug, and Cosmetic Act of 1938 that required materials sold to customers to be safe. In response to the need for test methods to assess ocular safety, *in vivo* animal tests were developed and put into use.

Although the rabbit was often used as the model in such testing, the methodology and species was not standardized until 1944, when FDA toxicologist J. H. Draize published a standard procedure for quantifying ocular injury (Draize et al., 1944). The resulting eye irritation procedure, also known as the Draize Test, generally involves applying 0.1ml or 0.1g of a

Received 27th July 2009, received in revised form and accepted for publication 28th February 2010

test substance to the eye of a restrained, conscious animal, for an exposure time of one to twenty-four hours. The animals are observed for up to 21 days for signs of corneal opacity, area of corneal opacity, iritis, conjunctival redness, edema, and discharge. Other effects, such as ulceration, hemorrhaging, cloudiness and vascularization are assessed in the tested eye. The test species is commonly an albino rabbit, though other species are used when required. The animals are humanely euthanized after testing is completed. The interested reader will find additional information on the historical aspects of the acute eye irritation test in several recent reviews (Parascandola, 1991; Wilhelmus, 2001; Wilson-Sanders, 2008).

2 The development of alternatives for the *in vivo* eye irritation test and early validation efforts

Over the years, there have been many criticisms of the Draize test, one of the most frequent being its use of animals. This concern provided the impetus for activists, academics, regulatory authorities and industrial scientists to find alternative methodologies that will provide reliable safety data without the need for *in vivo* tests. The need for alternative eye irritation tests has been particularly important to businesses that develop products requiring comprehensive safety testing before marketing to consumers who want products that are not tested in animals. The drive to eliminate animal testing has been particularly keen for companies that develop and market cosmetic and other types of personal care products.

It is not surprising therefore that the cosmetics and consumer products industries have invested heavily in the development and validation of non-animal replacements for the eye irritation test. Significant contributions by the industry include 1) research that led to the development of several novel non-animal tests, 2) involvement in validation studies from the mid-1990s, 3) sponsorship of key workshops where important developments were shared across the research community, 4) the conduct of basic mechanistic research supporting the development of trustworthy tests, and 5) data sharing. It is also important to note that other organizations have contributed significantly to the development effort. Academic institutions such as the Center for Alternatives to Animal Testing (CAAT, The Johns Hopkins University, http://caat.jhsph.edu/) and Fund for the Replacement of Animals in Medical Experiments (FRAME, University of Nottingham, UK) developed fundamental concepts and structural frameworks around which reliable tests could be developed. Government- and regulatory agency-sponsored organizations, such as the Interagency Coordinating Committee for the Validation of Alternative Methods (ICCVAM, USA), the German Bundesgesundheitsamt (ZEBET-BGA), and the European Center for the Validation of Alternative Methods (ECVAM, European Commission, http://ecvam.jrc.ec.europa.eu/), provided guidance on regulatory needs, developed practical validation processes and oversaw the effective execution of several key validation studies including:

1. European Commission / Home Office (EC/HO) study
2. European Cosmetic, Toiletry & Perfumery Association (COLIPA) study
3. Bundesgesundheitsamt / German Department of Research and Technology (BGA/BMBF) study
4. Cosmetics, Toiletry, and Fragrance Association (CTFA) study
5. Interagency Regulatory Alternatives Group (IRAG) study
6. Japanese Ministry of Health and Welfare / Japanese Cosmetic Industry Association (MHW/JCIA) study

Although these efforts did not identify any single assay as a full replacement of the animal test, they helped to define and refine the validation process (Balls et al., 1999; Bruner et al., 1996; Spielmann and Liebsch, 2001). They also led to discussion of the animal test and its reproducibility (Bruner et al., 1996; Prinsen, 2006), but regulatory use was not affected.

In 2003 the EU Cosmetics Directive was amended for the 7th time (European Union, 2003; DG ENTR, 2008; Hartung, 2008; Zuang et al., 2008). This amendment banned the use of the Draize eye test in Europe for finished cosmetic products from September 2004 and for cosmetic ingredients from March 2009. The anticipation of this ban led to major activities coordinated by two European Commission services, i.e. DG Enterprise and ECVAM (Hartung, 2008), especially compilation of an inventory of available methods (Eskes and Zuang, 2005). For eye irritation various assays including the CM were reviewed at a meeting coordinated by ECVAM in 2005 in order to identify the most promising tests for reducing and eventually replacing the animal test. Discussions at this meeting also led to the establishment of what is known as the Bottom-Up and Top-Down Approach for eye irritation testing (Scott et al., 2010).

The gathering of all available data made it possible for ECVAM to conduct retrospective validation activities for specified domains of applicability. A range of assays based on diverse mechanisms of action was evaluated (Hartung, 2007a). The objective of these exercises was to find specific ranges of irritation where a test could be considered valid, thereby providing an opportunity to combine one or more tests to cover the entire range of irritation response (Eskes et al., 2007; Scott et al., 2010; Zuang et al., 2008). Notably, the effort was strongly coordinated with ECVAM's US counterpart ICCVAM/NICEATM (The National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) and the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM), see http://iccvam.niehs.nih.gov/methods/ocutox/ivocutox.htm). The application of this retrospective validation approach to the CM will be described following a brief review of the steps that led to the development of the CM-based eye irritation test.

3 Introduction to the history of the Cytosensor Microphysiometer

For many years, the prospect of finding a replacement method for the Draize eye test was considered the Holy Grail of alternatives, much sought after but ever elusive. The quest led

a start-up company, Molecular Devices (Menlo Park, CA, USA), to consider the application of their novel CM technology for use in toxicity testing. They had published a paper in Science (McConnell et al., 1992) describing a technology that allowed the automated measurement of cellular metabolism in living cells. CAAT director Alan Goldberg, working with them as a consultant, guided the company to address alternatives to eye irritation. Alan Goldberg and researchers at Molecular Devices conducted a small pilot study that delivered promising results. With these findings in hand he forged a contact between Molecular Devices and The Procter & Gamble Company (P&G) that led to a collaboration establishing the first industrial laboratory to apply this instrumentation to the evaluation of eye toxicity.

The CM (Hafner, 2000) is based on the measurement of small changes in cellular metabolism reflected by extracellular release of acidic byproducts of energy metabolism. Energy metabolism in living cells is tightly coupled to cellular ATP usage and extracellular proton release. This means that events perturbing cellular ATP metabolism, such as receptor activation and initiation of signal transduction, will rapidly result in a change in the release rate of protons that can be measured in the culture medium surrounding cells. As the extrusion of protons is a very general indicator of cellular response to a broad range of perturbations, this endpoint can be used to assess the influence of chemicals on cells without prior knowledge of the corresponding signaling pathways. Extracellular acidification is measured in the CM using Molecular Device's proprietary technology called the light-addressable potentiometric sensor (LAPS).

In the early stages of the method's development P&G's Leon Bruner and Ron Parker met with developers of the Cytosensor Microphysiometer, Harden McConnell, Jack Owicki, and Wally Parce at Molecular Devices. They established a collaborative project to test a set of model eye irritants. Results from this pilot study appeared promising, leading P&G to bring a prototype instrument in-house for further investigations. More studies then followed which were published in 1991 (Bruner et al., 1991; Bruner and Parker, 1991). Following publication of these results, procedures were implemented within P&G to begin evaluating compatible test substances using the *in vitro* procedure instead of the rabbit-based eye irritation test.

At that time P&G also recognized that in order for an *in vitro* test to eventually become broadly accepted as a replacement for a traditional animal test, it would have to be commercially available to toxicologists from many companies and the regulatory community. In order to address this need, P&G approached Rodger Curren, head of a newly established *in vitro* toxicology division at Microbiological Associates (MA), and arrangements were made to transfer the technology from P&G to MA. Over the next few years the two companies worked together to refine test protocols and develop a larger database. Current standardized protocols are available from the INVITTOX database (protocol 97 and 102, http://ecvam-dbalm.jrc.ec.europa.eu/).

Other companies, for example L'Oréal, also began to publish data from the use of the instrument (Catroux et al., 1993). In 1997 Rodger Curren and others created a non-profit organization, the Institute for In Vitro Sciences, Inc. (IIVS, http://www.iivs.org/), which continued to work with the CM, especially for its capacity to differentiate degrees of mildness for cosmetic and personal care products, generally water-soluble surfactants and their formulations. P&G have remained a driving user of this method, and others, specifically L'Oréal and the Institute for In Vitro Sciences (IIVS), have also utilized the technology.

4 Validation of the CM for eye irritation testing

As noted above, the CM was included in some of the six major validation studies for eye irritation alternatives conducted from 1991-1997 including the European Commission / British Home Office (ECHO) study (Balls et al., 1995), a major study by the US Cosmetics, Toiletries and Fragrances Association (CTFA) (Bagley et al., 1992), and a study sponsored by the European Cosmetics Trade Association, Colipa (Brantom et al., 1997). Neither the CM nor any other of the methods evaluated were shown capable of fully replacing the animal test.

Essentially no new validation studies were undertaken until Thomas Hartung took over at the European Centre for the Validation of Alternative Methods (ECVAM) in 2002. Hartung opened validation to the principles of retrospective meta-analysis and weight-of-evidence approaches through the application of concepts from evidence-based medicine.

The next step was to use this approach to determine if an *in vitro* test could provide reliable information for safety assessments across the entire range of ocular irritancy responses. Concurrently, an advancement to the historic approach to validation gained consensus at a 2005 ECVAM Workshop led by Laurie Scott, a seconded expert to ECVAM from the Procter & Gamble, Co. that resulted in the identification of the Bottom-Up and Top-Down Approach to Eye Irritation (Scott et al., 2010). This approach introduced novel concepts to validation allowing methods to be assessed and validated for unique applicability and severity range. The analysis began with a review of the most promising assays for eye irritation. Based on guidance from the ECVAM Eye Irritation Task Force and other sources (Eskes et al., 2005, 2007; Scott et al., 2010; Zuang et al., 2008), ECVAM chose to assess the utility of the CM methodology, along with three other cell function and cytotoxicity methods, as a replacement for the eye irritation test. Valérie Zuang and Chantra Eskes at ECVAM were appointed leaders of the retrospective analysis.

This retrospective analysis was conducted by gathering and combining CM data from multiple prior studies in order to analyze them statistically using the modular approach (Hartung et al., 2004) and weight-of-evidence validation principles (Balls et al., 2006), and evaluation in accordance with the Bottom-up and Top-Down approach (Scott et al., 2010).The 2.5-year long retrospective validation effort allowed a compilation and analysis of all available data on the CM. The results were reported in a Background Review Document created by IIVS (available at http://ecvam.jrc.ec.europa.eu/) under a contract from ECVAM.

The analysis of the CM data showed that when the CM assay identified a surfactant or a surfactant-based mixture as being non-irritant, there was a high probability that the same material would be identified as a non-irritant by the Draize test. The analysis also showed that when the CM assay identified a water-soluble material or mixture as a severe irritant there was a high probability that the same material would be a severe irritant in the Draize test.

These results led to recommendations from the ECVAM Validation Management Group, following which, the CM proceeded to Peer Review by the ECVAM Scientific Advisory Committee, ESAC. The ESAC review was encouragingly positive. The ESAC not only accepted the VMG recommendations on the usefulness of the CM for the identification of severe eye irritants for water-soluble substances and mixtures in general, it also concluded positively on the usefulness of the CM for the identification of non-irritant water-soluble surfactants and water-soluble surfactant-containing mixtures (Box 1). ESAC also recommended that non water-soluble solids, suspensions or viscous material must be evaluated for

Box 1

ESAC Statement of validity for the Cytosensor/ Microphysiometer (extract)
(ecvam.jrc.it/publication/ESAC31_CBA_eye-irritation_20091005.pdf)

European Commission
Joint Research Centre
Institute for Health and Consumer Protection
In vitro methods Unit
European Centre for the Validation of Alternative Methods (ECVAM)

Statement on the Scientific validity of cytotoxicity/ cell function based *in vitro* assays for eye irritation testing

At its 31st meeting, held on 7 and 8 July, 2009 at the European Centre for the Validation of Alternative Methods (ECVAM), Ispra, Italy, the non-Commission members of the ECVAM Scientific Advisory Committee (ESAC) unanimously endorsed the following statement:

The replacement of traditional animal-based test methods by alternative ones should ideally be obtained by one-to-one replacements: to keep the testing regime simple and economical one single alternative method should, wherever feasible, be sufficient to generate data of equal or better quality than the traditional test.

However, in the case of eye irritation it is currently generally accepted that, in the foreseeable future, no single *in vitro* eye irritation test will be able to replace the *in vivo* Draize eye test to predict across the full range of irritation for different chemical classes. However, strategic combinations of several alternative test methods within a (tiered) testing strategy may be able to replace the Draize eye test.

A possible conceptual framework for such a (tiered) testing strategy has been developed within an ECVAM workshop (Ref. 1). The framework is based on alternative eye irritation methods that vary in their capacity to detect either severe irritant substances (EU R41; GHS 'Category 1') or substances considered non-irritant (EU 'Non-Classified'; GHS 'No Category').

According to this framework the entire range of irritancy may be resolved by arranging tests in a tiered strategy that may be operated from either end: to detect first severe irritants and resolve absence of irritancy ("Top-Down Approach") or to proceed inversely, starting with the identification of non-irritants first ("Bottom-Up Approach"). Mild irritancy will be resolved in a last tier in both approaches.

To evaluate the scientific validity of possible building blocks of such a test strategy and to assess their possible placement within a Bottom-Up and Top-Down Approach, ECVAM has undertaken a retrospective validation study of four cell-based *in vitro* methods. The test methods evaluated were:

a. Cytosensor Microphysiometer (INVITTOX Protocols 97 and 102 modified)2
b. Fluorescein Leakage (INVITTOX Protocols 71, 82, 86 and120);
c. Neutral Red Release (INVITTOX Protocol 54 and PREDISAFE™);
d. Red Blood Cell haemolysis (INVITTOX Protocols 37 and 99),

The four test methods, including ten protocol variations, were subjected to independent, expert review with respect to their use to either

a) initiate a Bottom-Up Approach, for consideration for regulatory use to identify non irritants (EU: 'Non Classified'; GSH: 'No Category'; EPA: 'Category IV') from all other classes as part of a tiered testing strategy, or
b) to initiate a Top-Down Approach, to identify ocular corrosives and severe irritants (EU R41, GHS 'Category 1', and EPA 'Category I') from all other classes as part of a tiered testing strategy.

In the absence of internationally agreed performance criteria for either approach, the PRP of the ESAC applied the following criteria:

– any test used to initiate a Top-Down Approach must balance specificity and sensitivity to correctly identify a substantial proportion of severe irritants, with a false positive rate that would not lead to the over-classification of an unreasonable number of materials of lower ocular irritancy potential – an

eye irritancy potential in other complementary *in vitro* test methods. This was the first time that an *in vitro* test had been considered accurate enough to accept its identification of a non-irritant, since previously *in vitro* methods had only been accepted to label a material as an irritant. Although the applicability domain for the CM is limited, the outcome represents a major breakthrough. This is the first time an assay has been endorsed as scientifically valid for the identification of non-irritant substances based on a retrospective weight-of-evidence evaluation. In addition, formal validation of the CM for the identification of both non-irritants and severe eye irritants allows its use in either a bottom-up or a top-down approach for eye irritation testing (Scott et al., 2010).

It is important to note that the CM test method cannot be considered a full-replacement method on its own because of misclassifications in the middle range of irritancy, i.e. irritation levels between severe or corrosive on one end, and non-irritating on the other. In this area there is a high false positive rate (non-irritants identified as irritants) and false negative rate (severe irritants identified as less than severe irritants).

over-classification rate (false positives) of <10% was considered acceptable
– any test used to initiate a Bottom-Up Approach should ideally give no false negatives with respect to human safety, and no false negative should be produced by high moderate or severe irritants.

Following independent ESAC peer review of this retrospective validation study and considering the potential test strategies in which the tests may be used, the ESAC concluded the following:

1. Cytosensor Microphysiometer Test Method
The Cytosensor Microphysiometer test method can be used for two of the three EU and GHS classification categories used for the endpoint of ocular irritation:

A. The **Cytosensor Microphysiometer test method (INVITTOX Protocol 102 modified)** is considered to have been scientifically validated and to be ready for consideration for regulatory use as an initial step within a **Top-Down Approach** to identify ocular corrosives and severe irritants (EU R41, GHS Category 1, and EPA Category I) from all other classes for the chemical applicability domain of water-soluble chemicals (substances and mixtures).

B. Furthermore, the **Cytosensor Microphysiometer test method (INVITTOX Protocol 102 modified)** is considered to have been scientifically validated and to be ready for consideration for regulatory use as an initial step within a **Bottom-Up Approach** to identify non-irritants (EU:NC; GHS: NC; EPA: cat IV) from all other classes only for water-soluble surfactants and water-soluble surfactant-containing mixtures.

C. On the basis of a thorough evaluation of the data compiled in the course of the ECVAM validation study, the ESAC concludes that the **Cytosensor Microphysiometer** test method does NOT correctly identify moderate and mild ocular irritants (EU: R36; GHS: Cat 2A/B; 85 EPA: Cat II/III). Therefore, the test method can only be employed to make decisions on two of the three categories of the eye irritation classification scheme (see A and B). Consequently, ESAC does NOT recommend this test method as a full replacement method. It should be noted in this context that the **Top-Down and Bottom-Up Approach** foresees the theoretical possibility of a *default* mild/moderate categorization (e.g. EU R36 or GHS Cat 2) of all those substances neither identified as ocular corrosives and severe irritants (see A) nor as "nonEUROPEAN COMMISSION classified" substances (see B) in the first two tiers of the strategy. However, the test method's high false negative rate (9-55%) when initiating a top-down approach and high false positive rate (50-69%) when initiating a bottom-up approach exclude the possibility to use the method for default categorization. The test methods can thus not be considered a full-replacement method on its own using the Top-Down and Bottom-Up approach.

Although these recommendations are based on the evaluation of data sets obtained using specific hard- and software, it is anticipated that other Cytosensor Microphysiometer equipment and software may become available with either equivalent or better performance and will need to be efficiently validated. Depending on the similarity of new equipment with respect to the validated one, this may be performed as a *Similar Method Validation* ('me-too') or an *Update Validation*. ESAC therefore recommends the development of Performance Standards for the Cytosensor Microphysiometer test method.

The current chemical applicability domain is limited: whilst in some cases this might be increased by expanding the data set of studied compounds, the test method is not amenable to testing non-water soluble solids, suspensions, or viscous materials.
[…]

Joachim Kreysa
Head of Unit
In vitro methods Unit
European Centre for the Validation
of Alternative Methods

Ispra, 10th July 2009

This does not interfere with its use to identify the extremes in either the Top-Down (for positive identification of severe irritants) or Bottom-Up approach (positive identification of non-irritant, i.e. not-classified materials). If one considers that ca. 80% of newly registered substances represent non-labeled substances, and ca. 15% represent severe eye irritants (Scott et al., 2010), the use of the CM in a Top-Down or Bottom-Up approach could overall allow for a significant decrease in animal testing for eye irritation.

It should be noted that this retrospective validation was undertaken even though Molecular Devices no longer manufactures or sells the instrument. However, it was ascertained before the evaluation began, that the CM construction plan is still available in the public domain, allowing its use by any interested party. Additionally, there are other instruments that operate on similar principles, which are now commercially available. If these other instruments are shown to be sufficiently similar to the validated CM system, a catch-up validation study could be used to demonstrate validity for testing similar domains of substances and irritancy ranges. A catch-up validation study is an abridged process for methods using similar instruments, software, and scientific principles. It requires agreed performance standards to confirm similarity. The new test method is required to provide similar or better accuracy and reliability compared to the existing validated method.

5 Evolution and history of validation concepts and their impact on the outcome and validation status of eye irritation methods

The field of eye irritation has been at the core of validation of alternative methods over more than two decades. Initially, a straight forward ring trial of novel methods comparing with historical data for broad chemical classes and all severity ranges was used as the default. The design was later termed "prospective", borrowing from terminology in epidemiology, since all data originate from experiments after the start of the validation exercise. Such design allows most control over the quality and composition of data and reduces many selection biases. As no scientific approach or regulatory guidelines existed for the experimental validation of *in vitro* toxicity tests, in 1990 a validation workshop with eminent scientists from both the EU and US sponsored by CAAT and ERGATT (European Research Group for Alternatives in Toxicity Testing) agreed in Amden, Switzerland on a simple definition of the validation process (Balls et al., 1990; Spielmann and Liebsch, 2001). Several international validation studies – especially in the field of eye irritation as described above – failed to identify one assay as a full replacement for the Draize eye test, although they were conducted according to these recommendations. Taking into account the lessons learned from this experience, a second validation workshop was held by ECVAM in Amden in 1994 (Balls et al., 1995) to develop a more precisely defined validation concept. Pre-validation (Curren et al., 1995), the development of biostatistically defined prediction models (Bruner et al., 1996), and a well-defined management structure were added as essential elements of the validation process. In 1995/1996 the ECVAM validation procedure was officially accepted by EU Member States and at the international level by the US regulatory agencies and the OECD. The improved validation concept was immediately introduced into ongoing validation studies and, in fact, the *in vitro* phototoxicity test was the first to successfully complete validation using these procedures.

This prospective design is still the default for the validation of any novel method. However, increasingly needs for adapting and making the process more flexible were voiced in the following years. Coming from the field of clinical pharmacology, Thomas Hartung as head of ECVAM brought in a new concept to the validation of alternative methods, i.e. to combine data from different studies - like a meta-analysis in clinical medicine (Hunt, 1997). The challenge is how to combine studies that were designed by different people, in different places at different times (Hartung, 2009). The opportunity, identified and then formalized in the modular approach to test validation (Hartung et al., 2004), was the introduction of retrospective validation. By putting forward the modular approach to validation, several amendments were introduced:

– The validation needs were defined as modules that can be filled independently by making use of existing data (retrospective analysis) or new prospective studies as well as combinations of both.
– The separation of the reproducibility and the relevance module allowed future consideration of new, lean designs of studies: while a ring trial is required to establish reproducibility requiring usually only a few substances, the broader testing of test substances for comparison with a reference test can be done in single laboratories.
– The test definition was made into a module of its own, acknowledging that definition of test SOP, prediction model and purpose is a key element of the validation process.
– The concept of applicability domains was adapted from *in silico* approaches, i.e. every test requires a clear statement on which test substances its validity has been established for.
– Performance standards were introduced to allow catch-up validation of similar methods with a lighter validation program.

The idea of retrospective validation was simple: Most tests entering validation have already been in use for a while – why not take into account all this information rather than start validation studies from scratch as if nothing were known? The term "retrospective" should distinguish this approach from prospective, new studies, borrowing terms from epidemiology and clinical study designs. The approach of retrospective validation has since 2004 been applied to areas such as eye irritation and has now led to the successful identification of a first method, the CM, to identify not only severe eye irritants but also substances not labeled as irritants. Noteworthy, the modular approach introduced also the concept of applicability domain, which has been much more strongly developed in the field of (Q)SAR and is less used in the field of *in vitro* toxicity testing. In the end the applicability domain concept has allowed the CM to be

considered valid for a very limited area of chemical substances or, the other way around, has restricted the use area of validity for the CM for the time being.

Validation is a continuously evolving field (Hoffmann and Hartung, 2006; Hartung, 2007b), shaped by accumulating experiences and emerging technologies. Frequently, barriers are encountered (Ahr et al., 2008), but this has regularly prompted adaptive responses, such as for intellectual property rights (Linge and Hartung, 2007), *in silico* methods (Hartung et al., 2004) or genomics approaches (Corvi et al., 2007). A key challenge is the point of reference, i.e. comparison to a gold (reference) standard. Two workshops addressed this area (Balls et al., 2006; Hoffmann et al., 2008), but without formal adaptations of new practices so far. The first workshop in 2004 (Balls et al., 2006) defined the principles of weight-of-evidence approaches to validation, where data could be analyzed and combined retrospectively. In conjunction, practical questions and approaches to advance the regulatory acceptance and use of *in vitro* methods validated for limited domains of applicability were developed during an ECVAM workshop in 2005 (Scott et al., 2010). The third workshop (Hoffmann et al., 2008) addressed the point of comparison in validation, i.e. what constitutes a good reference point. The eye irritation area would represent a key example to review the point of reference for validation, given the limitations of the Draize test as discussed earlier. Some efforts to have the Low Volume Eye Test (LVET) recognized as such a novel point of reference have been of limited success.

Over the last few years, discussion has moved in all areas of health hazards toward integrated testing strategies. It is both recognized that current (animal) tests require complementary approaches and that very few novel methods will individually replace our test needs one for one. Again, the area of topical toxicity was most advanced here already when the test guidance for REACH was developed and an integrated test strategy could be agreed upon. Going further, an ECVAM workshop held in 2005 (Scott et al., 2010) developed guidance for the integration of various alternative approaches for eye irritation in a Top-Down approach (identifying severe irritants first) and a Bottom-Up approach (identifying non-irritants first). Noteworthy, the ESAC statement sees a role for CM for both strategies in this testing approach.

6 Future needs and opportunities

The retrospective analysis of eye irritation methods has not yet fully leveraged the concept and principles of meta-analysis (Hartung, 2009), i.e. combining all available data and identifying influential variables. Instead, a core set of data produced according to a consensus protocol was distilled, which often left far too few data to identify applicability domains. Such analysis also needs weighing of data, e.g. as prepared for by developing a scoring tool for the quality of toxicological data (Schneider et al., 2009). The data-rich field of eye irritation could once again serve as a forerunner here.

The acceptance of the CM method as a first building block of a tiered testing strategy can now be expanded in two ways: The applicability domain of the CM might be expanded by providing additional evidence of valid results. This could be done either with the existing CM equipment, which is still used for toxicity testing, or with new equipment designed to measure cellular responses similar to those measured by the CM and which has been validated in a catch-up study. The second way forward is to populate the array of validated assays for eye irritation further in order to combine them with the CM in integrated testing strategies. Here still missing are other methods to compose and validate testing strategies systematically.

The limitations of the *in vivo* test as a reference standard, especially in the middle range of mild/moderate irritants (Bruner et al., 1996; Prinsen, 2006), remain a key challenge in the field. It is probable that an inclusion of the animal test in previous validation studies would likely have shown that the test is no better at reproducing itself than the *in vitro* alternatives (Bruner et al., 1996). Statistical approaches like latent class analysis (http://www.john-uebersax.com/) might offer some opportunities to validate new methods without comparing to the "gold (reference) standard" of the Draize test, but the regulatory acceptability of such approaches is not clear.

Furthermore, the field has not seen many contributions from the *in silico* field. These and other novel information-rich technologies, with the appropriate integration as systems toxicology (Hartung and Leist, 2008), might represent a way forward for eye toxicology in the 21st century.

7 Conclusions

More than 25 years of collaboration between various partners has led to the first validated method to identify non-irritant substances for the eye, albeit for a limited applicability domain. Remarkably, the Cytosensor Microphysiometer was a high-tech product at its time, commercialized by a university spin-off company. It was not originally developed to serve as an alternative method, but with the advice from a competence center in the field of alternatives this avenue was chosen. Major efforts especially by the cosmetic and consumer product industries conveyed the endurance to follow this approach. Notably, the product itself was discontinued some years ago, but the equipment/construction plans were made publicly available, a prerequisite for the validation. The overall duration of development, optimization and validation is one reason for the need for continued public funding and coordinated programs for alternatives.

The retrospective use of existing data, combining various studies, has again proven to be a powerful approach to extracting information from existing results. A consequent further development of such meta-analysis is urgently required, i.e. the development of processes, statistical tools, weighing scores, procedures and piloting examples. This represents a key contribution to evidence-based toxicology.

The broadening of the applicability domain and the addition of further methods in testing strategies is still needed and currently ongoing (Zuang et al., 2010; Freeman et al., 2010). However, the proof of principle that biotechnology approaches can replace rabbits in identifying non-irritant substances to the eye has been delivered. It might still be regarded by some to be only a small step in the quest to finally overcome the Draize rabbit eye test, but it is a big step for every rabbit spared…

References

Ahr, H.-J., Alepee, N., Breier, S. et al. (2008). Barriers to validation. A report by European Partnership for Alternative Approaches to animal testing (EPAA) working group 5. *ATLA 36*, 459-464.

Bagley, D. M., Bruner, L. H., de Silva, O. et al. (1992). An evaluation of five potential in vitro alternatives to the in vivo rabbit eye irritation test. *Toxicol. In Vitro 6*, 275-284.

Balls, M., Amcoff, P., Bremer, S. et al. (2006). The principles of weight of evidence validation of test methods and testing strategies. The report and recommendations of ECVAM workshop 58. *ATLA 34*, 603-620.

Balls, M., Berg, N., Bruner, L. H. et al. (1999). Eye Irritation Testing: The Way Forward. *ATLA 27*, 53-77.

Balls, M., Botham, P., Bruner, L. H. and Spielmann, H. (1995). The EC/HO International Validation Study on Alternatives to the Draize Eye Irritation Test. *Toxicol. In Vitro 9*, 871-929.

Balls, M., Botham, P. A., Fentem, J. H. et al. (1995). Practical aspects of the validation of toxicity test procedures. The report and recommendations of ECVAM workshop 5. *ATLA 23*, 129-147.

Balls, M., Blaauboer, B. and Brusik, D. (1990). Report and recommendations of the CAAT/ERGATT workshop on the validation of toxicity test procedures. *ATLA 18*, 313-337.

Brantom, P. G., Bruner, L. H., Chamberlain, M. et al. (1997). A summary report of the COLIPA international validation study on the alternatives to the Draize rabbit eye irritation test. *Toxicol. In Vitro 11*, 141-179.

Bruner, L. H., Carr, G. J., Chamberlain, M. and Curren, R. D. (1996). Validation of alternative methods for toxicity testing. *Toxicol. In Vitro 10*, 479-501.

Bruner, L. H., Kercso, K. M., Owicki, J. C. et al. (1991). Testing ocular irritancy in vitro with a cellular biosensor. *Toxicol. In Vitro 5*, 277-284.

Bruner, L. H. and Parker, R. D. (1991). Evaluation of seven in vitro alternatives for ocular safety testing. *Fund. Appl. Toxicol. 17*, 136-149.

Catroux, P., Rougier, A., Dossou, K. G. and Cottin, M. (1993). The Silicon Microphysiometer for Testing Ocular Toxicity In Vitro. *Toxicol. In Vitro 7*, 465-469.

Corvi, R., Ahr, H.-J., Albertini, S. et al. (2006). Validation of toxicogenomics-based test systems: ECVAM-ICCVAM/NICEATM considerations for regulatory use. *Environ. Health Persp. 114*, 420-429.

Curren, R. D., Southee, J. A., Spielmann, H. et al. (1995) The role of prevalidation in the development validation and acceptance of alternative methods. *ATLA 23*, 211-217.

DG ENTR (2008) Report from the Commission to the Council and the European Parliament – Report on the Development, Validation and Legal Acceptance of Alternative Methods to Animal Tests in the Field of Cosmetics (2007) – COM(2008) 416 final. http://ec.europa.eu/enterprise/cosmetics/html/cosm_animal_test.htm (last accessed 19 August 2009).

Draize, J. H., Woodard, G. and Calvery, H. O. (1944). Methods for the study of irritation and toxicity of substances applied topically to the skin and mucous membranes. *J. Pharmacol. Exp. Therapeut. 82*, 377-390.

Eskes, C., Zuang, V. and Hartung, T. (2007). The Way Forward for Eye Irritation Testing. AltTox website: http://www.alttox.org/ttrc/toxicity-tests/eye-irritation/way-forward/eskes-zuang-hartung (accessed 09.04.2009).

Eskes, C., Bessou, S., Bruner, L. et al. (2005). Subchapter 3.3. Eye Irritation. In Alternative (non-animal) Methods for Cosmetics Testing: Current Status and Future Prospects (Eskes C., Zuang V. eds). *ATLA 33, Suppl. 1*, 47-81.

European Union (2003). EC – directive 2003/15/EC of the European parliament and of the council of 27 February 2003 amending council directive 76/768/EEC on the approximation of the laws of the member states relating to cosmetic products. Official Journal L66, 11/03/2003, 26-35. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2003:066:0026:0035:en:PDF (last accessed 24 February 2010).

Freeman, S. J., Alépée, N., Barroso, J. et al. (2010). Prospective validation study of reconstructed human tissue models for eye irritation testing. *ALTEX 27, Suppl. Proceedings WC7 Rome*, in press.

Hafner, F. (2000). Cytosensor® Microphysiometer: technology and recent applications. *Biosensors Bioelectron. 15*, 149-158.

Hartung, T. (2009). Food for thought … on evidence-based toxicology. A*LTEX 26*, 75-82.

Hartung, T. (2008). Food for thought … on alternative methods for cosmetics safety testing. *ALTEX 25*, 147-162.

Hartung, T. (2007a). ESAC statement on the conclusions of the ICCVAM retrospective study on organotypic in vitro assays as screening tests to identify potential ocular corrosives and severe irritants as determined by US EPA, EU (R41) and UN GHS classifications in a tiered testiung strategy, as part of a weight of evidence approach. *ATLA 35*, 305.

Hartung, T. (2007b). Food for thought … on validation. *ALTEX 24*, 67-72.

Hartung, T. and Leist, M. (2008). Food for thought … on the evolution of toxicology and phasing out of animal testing. *ALTEX 25*, 91-96.

Hartung, T, Bremer, S., Casati, S. et al. (2004). A modular approach to the ECVAM principles on test validity. *ATLA 32*, 467-472.

Hoffmann, S., Edler, L., Gardner, I. et al. (2008). Points of reference in validation – the report and recommendations of ECVAM Workshop. *ATLA 36*, 343-352.

Hoffmann, S. and Hartung, T. (2006). Towards an evidence-based toxicology. *Human Exp. Toxicol. 25*, 497-513.

Hunt, M. M. (1997). *How Science Takes Stock: The Story of Meta-Analysis* (1st edition). New York: Russell Sage Foundation Publications.

Linge, J. and Hartung, T. (2007). ECVAM's approach to intellectual property rights in the validation of alternative methods. *ATLA 35*, 441-446.

McConnell, H. M., Owicki, J. C., Parce, J. W. et al. (1992). The cytosensor microphysiometer: biological applications of silicon technology. *Science 257*, 1906-1912.

Parascandola, J. (1991). The development of the Draize test for eye toxicity. *Pharmacy in History 33*, 111-117.

Prinsen, M. K. (2006). The Draize eye test and in vitro alternatives: a left-handed marriage? *Toxicol. In Vitro 20*, 78-81.

Schneider, K., Schwarz, M., Burkholder, I. et al. (2009). "ToxRTool", a new tool to assess the reliability of toxicological data. *Toxicol. Lett. 189*, 138-144.

Scott, L., Eskes, C., Hoffmann, S. et al. (2010). A proposed eye irritation testing strategy to reduce and replace in vivo studies using bottom-up and top-down approaches. *Toxicol. In Vitro, 24*, 1-9.

Spielmann, H. and Liebsch, M. (2001). Lessons learned from validation of in vitro toxicity test: from failure to acceptance into regulatory practice. *Toxicol. In Vitro 15*, 585-590.

Wilhelmus, K. R. (2001). The Draize eye test. *Survey Ophthalmol. 45*, 493-515.

Wilson-Sanders, S. E. (2008). Research animal methods VSC/443/543 – Class notes for 10/01/08: Mrs. Brown's Sad Story: A History of the Food, Drug, and Cosmetic Act. http://www.uac.arizona.edu/VSC443/Alternmethod/Fdapap03.htm (last accessed 19 August 2009)

Zuang, V., Barroso, J., Cole, T. et al. (2010). ECVAM Bottom-Up/Top-Down Testing approach: testing strategy to reduce/replace the Draize eye test and validation/regulatory acceptance of in vitro assays: current status. *ALTEX 27, Suppl. Proceedings WC7 Rome*, in press.

Zuang, V., Eskes, C., Griesinger, C. and Hartung, T. (2008). ECVAM key area topical toxicity: update on activities. *ALTEX 14*, 523-528.

Acknowledgement

Success has many parents. The co-authors of this article, brought together by Thomas Hartung to accompany the publication of the ESAC statement of validity, gratefully acknowledge their contributions especially our colleagues from L'Oréal, even if not always named in person.

Correspondence to

Thomas Hartung, MD PhD
Doerenkamp-Zbinden Professor and
Chair for Evidence-based Toxicology
Johns Hopkins University
Bloomberg School of Public Health
Department of Environmental Health Sciences
Center for Alternatives to Animal Testing (CAAT)
615 N. Wolfe St.
Baltimore, MD, 21205, USA
e-mail: THartung@jhsph.edu