14. The Meeting acknowledged the Solna criteria for validation and regulatory acceptance are essentially the same as the ICCVAM criteria. In the case of percutaneous absorption the *in vivo* and the *ex vivo* Guidelines comprise methods with a long history and a database of varying quality,. Therefore, the Extended Steering Committee agreed that the Solna criteria should be considered with some flexibility. In the case of non-compliance, however, appropriate justification should be provided.

COLIPA Data

15. The COLIPA data package had been sent to the Nominated Expert Reviewers with copies to the Steering Committee on 4th August with the request to critically review the data. When it became obvious that members of the Steering Committee considered the data to be inadequate for the purpose of the Workshop, the Expert Reviewers were informed of this on 4th September. Some Expert Reviewers nonetheless appeared willing to continue with the review and had submitted their findings to the Secretariat. Most significant findings were:

- there were too many variables to allow for a comprehensive evaluation of any of the possible approaches;
- there was a lack of individual data, needed for a full evaluation;
- the study protocols provided were not sufficiently detailed;
- no information was provided on the criteria that were used to select the chemicals or methods for which data were provided;
- it was not always clear whether absorption values were those of receptor fluid only or also included the skin; there was in general a lack of detail of the analytical methods applied;
- however, those individuals who use the ex-vivo methods regularly indicated that they observed good agreement between in vivo and ex vivo data.]

16. Following the brief COLIPA data review, there was an extensive discussion of data submission guidelines. The Canadian/US proposal which had been circulated and discussed earlier by the Steering Committee was considered too rigid and too detailed by some (Extended) Steering Committee Members. Others felt that, although the Canada/US Data Submission Form is very detailed indeed, this should not automatically mean that it be used as a checklist, leading to rejection of a submission if some requested information is lacking. An alternative, one page, table format, data submission evaluation form was presented by Jon Heylings, who had in fact used it for the evaluation of the COLIPA data. A considerable number of (Extended) Steering Committee Members appreciated this compact format, covering all elements essential for the evaluation of a validation study. Others, however, were more skeptical and were of the opinion that its use could lead to an overly liberal acceptance of new tests without providing sufficient confidence in their reliability and relevance. The Meeting finally came to the conclusion that the Canadian/US Data Submission Form is a very helpful tool for the evaluation of validation studies but that it should be used as guidance, rather than as a submission form. The one page form designed by Jon Heylings was also considered a good tool for this purpose, especially for a first assessment. The US/Canada proposal and Jon Heylings' form are attached to this report as Annex 3 and 4, respectively.

17. The Meeting agreed that the COLIPA data package could be further upgraded. Major improvements could be achieved by: (i) providing the criteria used for the selection of the chemicals and the studies presented, (ii) describing the various test protocols used in more detail in order to allow better comparisons, (iii) addressing the specific questions, raised by reviewers who continued their review after it became clear that the Workshop was canceled, and, (iv) using the proposal from Canada/US for a Submission Form, Appendices 1-4, together with Jon Heylings' one-page evaluation form, as guidance for data submission.

Other Data

- the Secretariat will make an effort to obtain additional industry data on in vivo/ex vivo comparisons of percutaneous absorption studies, especially for pesticides and pharmaceuticals;

- a Review Panel will be established to carry out an in-depth review of all available data sets; the Review Panel will report to the Steering Committee;

- a Guidance Document will be prepared which will be referred to in the Test Guidelines for percutaneous absorption and which will provide more detailed guidance on the use of the *exvivo* and the *in vivo* tests in risk assessment and will also provide more detail on preferred techniques in given situations.

23. The Meeting briefly discussed the options for follow-up after the Expert Panel Review Report becomes available. These options include:

i) a Workshop to discuss the outcome of the data review and the revised Guideline proposals. A Workshop would allow a broad participation of experts with emphasis on the discussions of the scientific aspects of percutaneous absorption testing and assessment; a Workshop is normally followed by a proposal which will be circulated to Member countries' experts for comments. Next, National Positions are requested and finally one or more Test Guidelines may be approved.

ii) a Nominated Expert Meeting to agree on the (non)acceptance of the proposed test Guidelines, amended as appropriate during the Meeting. A Nominated Expert Meeting comprises experts, nominated by their country to represent their country and its national position on the issue. These meetings are normally smaller in size than a Workshop. Nominated Expert Meetings are usually followed by a written procedure among National Co-ordinators of the Test Guidelines Programme to achieve final approval of the Test Guideline(s).

iii) Circulation for review of the revised Guideline proposals together with the report of the Expert Review Panel. This option would be appropriate if the report of the Review Panel provides an unequivocal conclusion about the available data and the consequences thereof for the Guideline proposals. In that case there will be no need for further extensive discussions and National Co-ordinators (and National Experts) will be provided with the appropriate proposals and asked to come to a final conclusion.

24. The Extended Steering Committee decided that soon after the report of the Expert review Panel becomes available, a meeting of the Committee should be arranged to agree on the appropriate option for further steps.

25. Details of the action plan/additional work and an indication of the time needed for the various activities are provided in Annex 5 and 6, respectively.

Annex 2

List of Participants of the Extended Steering Committee Meeting

Steering Committee Members:

ALLEMAGNE	
Mr. Walter DIEMBECK FZ-4232 BV/BC Beiersdorf AG Unnastrasse 48, Hamburg	49-40-4909 2662 49-40-4909 3589
Mr Hermann-Georg HOLZHÜTTER Humboldt University of Berlin Medizinische Fakultat (Charite) Bereich Medizin Institute for Biochemistry Monbijoustr.2a, D-10117 Berlin	49-30 28 02 6391 49-30 28 02 6615
Mr Horst SPIELMANN Director und Professor ZEBET Diedersdorfer Weg 1 D-12277 Berlin	49-30 8412 2270 49-30 8412 2958 zebet@bgvv.de
CANADA	
Dr Ih CHU Head Environmental Contaminants Section Environmental Health Centre Tunney's Pasture Ottawa K1A OL2	613-957-1837 1-613 941 4768 Ih_Chu@hc-sc.gc.ca
Mr. John WORGAN Head, Occupational Exposure Asses. Section Pest Management Regulatory Agency Sir Charles Tupper Building, 2250 Riverside Drive, Address Locator 6605D Ottawa Ontario K1A OK9	1 613 736 3485 1 613 736 3489 jworgan@pmra.hwc.ca
ETATS UNIS	
Ms. Katherine A. STITZEL Associate Director The Proctor & Gamble Company Human Safety Miami Valley Laboratories 1180 East Miami River Road Ross, Ohio 45061	Tel.: 1-513-627-2965 Fax: 1-513-627-2188 stitzel.ka@pg.com
Mr. Robert L. BRONAUGH	1-301-594-5813

ROYAUME-UNI	
Mr. D. HOWES Unilever ESL Colworth House Shambrook Bedford	44-1234-781-781 44-1234-222-122
EUROPEAN COMMISSION	
Ms. Julia FENTEM ECVAM JRC Environment Institute Via Enrico Fermi 21020 Ispra (Varese)	39-32 332-789-036 39-32 332-785-336 julia.fentem@ei.jrc.it

Additionally Invited Experts:[1]

Mr. Neil CARMICHAEL Directeur Rhone Poulenc Agro Centre de Recherche 355 rue Dostoievski - B.P. 153 F-06903 Sophia Antipolis Cedex FRANCE	04 92 94 34 02 04 93 65 41 39
Mr. Jon HEYLINGS ZENECA Central Toxicology Laboratory Alderley Park Macclesfield Cheshire ROYAUME-UNI	44 1625 514 550 44 1625 586 396 E-mail : Jon.Heylings@ CTL.zeneca.com
Mr. Bernard SCHWETZ Director Research NCTR, FDA Nat. Centre for Toxological Research Building 13, Room 123C (HFT-1) NCTR Drive,Jefferson USA	1 501 543 7517 1 501 543 7576

Secretariat:

Mr. Herman B.W.M. Koeter OECD Environmental Health & Safety Division 2, rue Andre Pascal 75775 Paris Cedex 16	01 45 24 9844 01 45 24 16 75 E-mail:herman.koeter@oecd.org

[1] Dr. John Frazier (Senior Scientist, US AirForce) and Dr. Canice Nolan (EC.DG12) were also invited as additional experts but were unable toattend

Annex 3

OECD IN VITRO DERMAL PENETRATION TEST GUIDELINE:

ISSUE RESOLUTION PROCESS

(DRAFT, 23 June 1997)

A Canadian - U.S. Proposal

Cathleen Campbell
Ih Chu
John Worgan

Health Canada

Richard Hill

U.S. Environmental Protection Agency

William Stokes

National Institute of Environmental Health Sciences

ISSUES:

A draft OECD test guideline for Dermal Delivery and Percutaneous Absorption: In Vitro Method was distributed to member states for comment in June 1996. Most authorities thought the method was ready for acceptance. Canada and the U.S., however, questioned whether there was adequate documentation of the validation of the method. More specifically:

1. The purpose of the guideline is not articulated (e.g., screen, adjunct, replacement).

2. The applicability of the test for specific chemicals/classes of chemicals/universe of chemicals is not documented.

3. The applicability of the test for chemicals with differing water and lipid solubility is not provided.

4. The guideline is not a protocol and needs more specificity concerning the influence of the following variables on test outcome:

d. Establish data analysis procedures (attachment 3)

2. Evaluation activities

a. Solicit unpublished and published data in accordance with 1a and 1b above
b. Develop questions to be answered by data reviewers (attachment 4)
c .Conduct data evaluation
d. Apply criteria for validation and acceptance (as developed by OECD (attachment 5) and expanded upon by ECVAM and ICCVAM)
e. Write OECD test guidelines for methods that are valid and acceptable for specific uses

(Draft, 19 June 1997)

TEST METHOD SUBMISSION GUIDELINES (Summary)

INTRODUCTION AND RATIONALE

1. Indicate the scientific and regulatory rationale for the test method.

2. Describe the mechanistic and/or empirical basis of the proposed test.

3. Discuss the intended uses of the test method.

4. Indicate, as appropriate, if and how the proposed method can replace current tests.

5. Describe, if applicable, how the proposed method may reduce, refine, or replace animal use.

6. Give the cost and time requirements relative to current procedures.

TEST METHOD PROTOCOL

1. Provide the detailed protocol for the test method, including a complete description of the test model or substrate, duration of exposure, known limits of use, and nature of the response assessed.

2. Describe the nature of the data to be collected, including the endpoints measured and observations recorded, measures of variability, dose-response measurement, number of replicate studies.

3. Summarize control data, including number of trials, measures of central tendency and variability, etc.

CHARACTERIZATION OF MATERIALS TESTED

1. Describe the chemicals/products evaluated:

1. Provide copies of all relevant publications.

2. Summarize and provide the results of any peer reviews conducted to date, and summarize any other ongoing or planned reviews.

3. Discuss the availability of laboratory notebooks for audit

(Draft, January 1997)

SCREENING CRITERIA FOR EXISTING *IN VITRO-IN VIVO* STUDIES

DATABASE

All data sets submitted or collected from the literature will have specific information recorded in a bibliographic database (Endnote II): author, title, date, publication details and chemical name and whether or note they have met the screening criteria. Studies not meeting the screening criteria will not be considered further, however will be kept on record in the bibliography.

SCREENING CRITERIA

1. Animal species must be identified for both the *in vivo* and *in vitro* assays. If different animal species are used in the *in* vivo and *in vitro* assays, they will be included in the database.

2. Acceptable total mass balance is reported for both the *in vitro* and *in vivo* assays (acceptable total mass balance is defined as approximately ≥80%)

3. For both assays, mass balance is compartmentalized into:

   a. amount washed from the skin,
   b. amount retained at the skin site, and
   c. amount absorbed.

   How these compartments are to be used in absorption calculations will be defined at a later time.

4. Dose(s) for both the *in vitro* and *in vivo* assay are similar (i.e., within 5 fold).

5. Dose preparations (e.g., vehicle) are the same for both the *in vitro* and the *in vivo* assay. If similar, but not the same, vehicles (e.g. solvents) are used, they will be included in the database.

6. The dosing regimes are similar (e.g., both *in vitro* and *in vivo* studies have single dose applications).

It is important to note that dermal absorption is not a linear process either with exposure period or dose levels, therefore it will be necessary to examine *in vivo* and *in vitro* dermal absorption results across a range of dose and exposure periods.
At each dose level, the relationship between *in vivo* and *in vitro* data should be examined over all exposure periods. Any changes/trends in the *in vivo:in vitro* relationship over time may then be established.

At each exposure period, the relationship between the *in vivo* and *in vitro* data should also be examined over all dose levels. Analysis of this data may establish any changes in the *in vivo:in vitro* relationship over dose levels. In both analyses, consistent differences between methodologies (i.e. consistently under/ overestimating the *in vivo* results) may prove to be useful.

2. Inter-study Analysis (Meta-Analysis)

The comparative *in vivo:in vitro* results of a subset of studies representing specific methodologies may be presented to assess methodological effects. These sets of comparative data may be further subset (e.g. chemical classes with specific methodological protocols) to establish acceptable ranges of ratios of *in vivo* to *in vitro* dermal penetration results.

1. Comment on the adequacy of the statistical methods used to evaluate the performance of the test method.

2. Comment on the quality and acceptability of the in vivo reference data used to compare the performance of the in vitro method.

3. Comment on the adequacy of the chemicals/products (numbers/types) selected to evaluate the performance of the method for each chemical/product class. Is it appropriate to generalize the performance of the method for all chemicals/products in each class based on the performance of the selected test chemicals/products?

4. Are sufficient data provided to adequately evaluate the performance of the method for its proposed use?

5. Comment on the sensitivity, specificity, concordance, and frequency and magnitude of overprediction and underprediction for the specified exposure durations, dose ranges, and chemical/product classes that the method is proposed to be used for.
   (a) Does the method consistently predict the magnitude of dermal absorption for some or all chemical/product classes over the specified dose ranges and for the specified duration of exposure?
   (b) Does the method consistently predict the lack of dermal absorption for some or all applicable chemical/product classes over the specified dose ranges and for the specified duration of exposure?
   (c) Does the method consistently over or underpredict dermal absorption compared with the reference test method?

6. Does the method adequately predict the endpoint of interest by demonstrating a linkage between the test and the current in vivo test method?

7. Are the conclusions on the usefulness of this method scientifically sound?

DETERMINATION OF TEST METHOD RELIABILITY
(Repeatability and Reproducibility)

1. Comment on the adequacy of the evaluation of intralaboratory repeatability and reproducibility of the test method, and the data used to define and describe the level of intra-laboratory variability.

2. Comment on the adequacy of the evaluation of interlaboratory reproducibility of the test method, and the data used to define and describe the level of interlaboratory variation.

3. Was the test method's reproducibility evaluated on a series of appropriate reference chemicals or products, and do these adequately represent the types of substances for which the test method is proposed to be used?

OTHER SCIENTIFIC REVIEWS

Attachment 5

(Draft, 20 June 1997)

OECD VALIDATION AND ACCEPTANCE CRITERIA FOR TOXICOLOGICAL TEST METHODS

VALIDATION CRITERIA

Criteria that should be met for new animal or non-animal methods to be considered validated are:

1. A rationale for the test method should be available, including a clear statement of scientific need and regulatory purpose.

2. The relationship of the endpoint(s) determined by the test method to the in vivo biological effect and to the toxicity of interest must be addressed, and limitations of the method must be described.

3. A formal detailed protocol must be provided and should be readily available in the public domain. It should be sufficiently detailed to enable replication by other users, and should include data analysis and decision criteria. Test methods and results should be available, preferably in a peer reviewed publication, and test results should have been subjected to independent scientific review

4. Intra-test variability, repeatability and reproducibility of the test method within and amongst laboratories should have been demonstrated. Data should be provided describing the level of intra- and inter- laboratory variability and how these vary with time.

5. The test method's performance must have been demonstrated using a series of reference chemicals, preferably coded to exclude bias.

6. The performance of test methods should have been evaluated in relation to existing relevant toxicity data as well as information from the relevant target species.

7. All data supporting the assessment of the validity of the test methods, including the full data set collected in the validation study, must be available for review.

8. Normally, these data should have been obtained in accordance with OECD Principles of Good Laboratory Practice.

ANNEX 4

NAME OF COMPANY

Questions to peer reviewers

Test Method Description	1. In vitro / In vivo, species comparison	2. Chemical or formulation	3. Skin preparation	4. Objective and intended use of data
1a Protocol description				
1b How data to be used				
1c Decision criteria				
2 Skin Integrity				
3 Mass balance				
4 Limitations of the test				
Test Method Data Quality				
1 Adherence to protocol				
2 Changes to protocol				
3 GLP				
4 QA audit conducted				
Acceptance Criteria				
1 Linkage of new test to old test				
2 Use in risk assessment				
3 Adequate data for such chemicals				
4 Transferability				
5 Cost effective				
6 Justification e.g. 3Rs				

ANNEX 5

ACTION PLAN/ADDITIONAL WORK ON PERCUTANEOUS ABSORPTION

ACTIVITY	Who?
1. REVISION OF THE TEST GUIDELINES *(IN VIVO AND EX VIVO)* TO: • describe in more detail the intended use, including a section on "limitations and possibilities"; • narrow the range of technical options; • provide more guidance on specific details; • provide preferred options (choice of vehicle, rinsing, skin type, etc.); • consider exclusion criteria (caustic/corrosive substances, species not to be used)	Doug Howes and Bob Bronaugh
2. SELECTION OF REFERENCE CHEMICALS • identification of a set of chemicals covering the range of physico-chemical characteristics considered essential to evaluate the validity of the proposed test methods and which could be recommended for inclusion in the test as positive/negative controls.	Neil Carmichael, Ih Chu, Bob Bronaugh and Hans Schäfer.
3. LITERATURE SEARCH • Identification of the most relevant literature published since 1982-1983 on *ex vivo/in vivo* comparisons: a. Literature searches to collect a comprehensive list of available literature; b. Merging of the searches into a single exhaustive list; c. Identification of relevant articles by the application of objective criteria based on Jon Heyling's approach and the Canadian/US Proposal (Appendices 1-4) using abstracts or full papers (to reduce the exhaustive list to a shortlist of 25-100 studies); d. Request for individual data on a "need be" basis from the authors of the selected short list of publications; e. In depth review of the selected papers by the Review Panel (4-6 experts) using the Canadian/US Proposal (Appendix 1-4). Studies for which individual data cannot be provided will not be *a priori* rejected.	Sources: Alan Goldberg (awac); Hans Schäfer; Ih Chu (Canadian list); Jon Heylings; ECVAM list; ECETOC Monogr 20; Bob Bronaugh. 3a/b: Secretariat 3c: Jon Heylings and Ih Chu 3d: Secretariat 3e: Review Panel (to be established by the Steering Committee)
4. INDUSTRY DATA • Industry will be requested to provide data on *in vivo/ex vivo* comparisons for pesticides, pharmaceuticals and other chemical use categories, followed by: a. Screening of these additional data sets: accept or reject; b. In depth review of the accepted data sets using the Canadian/US Proposal (Appendices 1-4).	Secretariat; BIAC: ECPA and ACPA Task Forces 4a/b: Review Panel

Annex 6

1998 Time Schedule for additional work on percutaneous absorbption data collection and evaluation.

July '98	August	September	October	November	December	January	February	March
Revision of the Guidelines								
Reference Chemicals Selection								
Literature Search and Initial Evaluation								
			Individual Data Retrieval from Published Literature					
		Additional Industry Data Collection						
Further upgrading of Colipa Data Package								
					In depth Data Review - RPM[1] - SCM[2]			Follow-up Actions
Development of a Guidance Document								
NB Dates have changed								

1) RPM= Review Panel Meeting 2) Steering Committee Meeting

資料番号-10

First draft 7/12/98

Eye Irritation Testing: The Way Forward

The Report and Recommendations of ECVAM Workshop 32[1,2]

Michael Balls,[3] Ninna Berg,[4] Leon H. Bruner,[5] Rodger D. Curren,[6] Odile de Silva,[7] Lesley K. Earl,[8] David J. Esdaile,[9] Julia H. Fentem,[3] Manfred Liebsch,[10] Yasuo Ohno,[11] Menk K. Prinsen,[12] Horst Spielmann[10] and Andrew P. Worth[3]

[3]ECVAM, JRC Environment Institute, 21020 Ispra (VA), Italy; [4]Novo Nordisk A/S, Novo Alle, 2880 Bagsvaerd, Denmark; [5]The Procter & Gamble Company (Health and Beauty Care) Ltd, Lovett House, Lovett Road, Staines, Middlesex TW18 3AZ, UK; [6]Insititute for In vitro Sciences, 21 Firstfield Road, Gaithersburg, MD 20878, USA; [7]L'Oréal Recherche Avancée - Sciences du Vivant, 1 Avenue Eugène Schueller, 93601 Aulnay-sous-Bois Cedex, France; [8]SEAC Toxicology Unit, Unilever Research, Colworth House, Sharnbrook, Bedford MK44 1LQ, UK; [9]Rhône-Poulenc Agro, 355 rue Dostoievski, 06903 Sophia Antipolis Cedex, France; [10]ZEBET, BgVV, Diedersdorfer Weg 1, 12277 Berlin, Germany; [11]Division of Pharmacology, Biological Safety Research Centre, National Institute of Health Sciences, 1-18-1 Kamiyoga, Setakaya-ku, Tokyo 158, Japan; [12]TNO Nutrition and Food Research Institute, Division of Toxicology, 3700 AJ Zeist, The Netherlands

Preface

This is the report of the thirty-second of a series of workshops organised by the European Centre for the Validation of Alternative Methods (ECVAM). ECVAM's main goal, as defined in 1993 by its Scientific Advisory Committee, is to promote the scientific and regulatory acceptance of alternative methods which are of importance to the biosciences and which reduce, refine or replace the use of laboratory animals. One of the first priorities set by ECVAM was the implementation of procedures which would enable it to become well-informed about the state-of-the-art of non-animal test development and validation, and the potential for the possible incorporation of alternative tests into regulatory procedures. It was decided that this would be best achieved by the organisation of ECVAM workshops on specific topics, at which small groups of invited experts would review the current status of various types of *in vitro* tests and their potential uses, and make recommendations about the best ways forward (1).

The workshop on Eye Irritation Testing: the Way Forward was held in Egham, UK, on 15-17 June 1998, under the chairmanship of Michael Balls (ECVAM, Italy). The principal aims of the workshop were to: a) assess the reasons why many promising alternatives to the Draize eye irritation test have not been successful in multi-laboratory validation studies; b) discuss a new approach to the validation of *in vitro* tests for eye irritancy, based on the use of reference standards, which promises to overcome some of the problems encountered in

Address for correspondence and reprints: Professor Michael Balls, ECVAM, TP 580, JRC Institute for Health & Consumer Protection, 21020 Ispra (VA), Italy.

[1]ECVAM - European Centre for the Validation of Alternative Methods.
[2]This document represents the agreed report of the participants as individual scientists.

previous studies; and c) discuss new approaches for assessing eye irritation, such as the use of stepwise testing strategies and multivariate statistics.

Introduction

The Draize rabbit test (Draize *et al.*, 1944) continues to be the method of choice for the regulatory assessment of eye irritation hazard (OECD, 1992; EC, 1993), despite the fact it has been heavily criticised on both scientific and animal welfare grounds. This is not due to a shortage of potentially useful alternative methods, since more effort has probably been put into the development of alternatives to the Draize eye irritation test than into seeking replacements for all the other acute *in vivo* toxicity tests put together. However, no test, combination of tests, or testing strategy has yet been developed which meets all the requirements of the regulatory authorities. This is largely because of the nature of the *in vivo* test itself, which, being based on the subjective scoring of tissue lesions in the eye, provides highly variable estimates of eye irritancy (Earl *et al.*, 1997) which are difficult to compare in a meaningful way with *in vitro* test results. Thus, although there is much confidence that a number of the alternative tests and testing strategies do work in-house, it has proved impossible to establish this satisfactorily by conducting validation studies in which *in vitro* test results are compared with historical *in vivo* data (Balls *et al.*, 1995; Brantom *et al.*, 1997). To find possible solutions to this impasse, ECVAM organised a workshop on eye irritation testing, which brought together experts in the field from industrial and governmental organisations. The workshop participants reviewed the eye irritation validation studies carried out to date, and discussed the possible use of reference standards (benchmark chemicals) in the validation process, since this could overcome some of the problems which were encountered in previous studies. In addition, the possibilities offered by stepwise testing strategies, and multivariate statistics were discussed, and an approach for evaluating stepwise testing strategies was presented. This report summarises the workshop discussions, and presents the conclusions and recommendations of the workshop participants.

Review of Validation Studies

The EC/HO study

The validation study which has become known as the European Commission/British Home Office (EC/HO) study (Balls *et al.*, 1995) was set up in the light of an EC-funded pilot study (EC, 1991) to establish whether one or more of nine tests could be used to: a) replace the Draize test for all severely irritating substances or for severely irritating substances belonging to specific chemical classes; and b) replace the Draize test for all levels of eye irritancy with or without regard to the chemical class. The nine tests comprised: a) four cell culture methods (based on red blood cell [RBC] haemolysis, neutral red uptake [NRU], fluorescein leakage [FL], and the use of the silicon microphysiometer [SM]); b) three *ex vivo* tests (involving the isolated rabbit eye [IRE], the isolated chicken eye [ICE], and the isolated bovine cornea [BCOP]); c) the hen's egg test on the chorioallantoic membrane (HET-CAM); and a physicochemical method based on protein precipitation (EYTEX^{TM}).

The relevance and reliability of the nine test methods were assessed under blind conditions by using a test set of 60 single chemicals, which were independently selected and

coded/supplied to 37 laboratories. The data generated by the laboratories was analysed independently. The reliability of each test was assessed by determining the interlaboratory Pearson correlation coefficients of the *in vitro* scores for each endpoint. These analyses indicated that there was good reproducibility between the laboratories conducting the same test. The relevance of each test was assessed by: a) calculating the Pearson and Spearman correlation coefficients for the relationship between each alternative test endpoint and the Modified Maximum Average Score (MMAS); and b) deriving a simple linear regression equation to predict the MMAS from each alternative test endpoint and to determine a 95% confidence interval (CI) for this prediction. These analyses showed that, for the full set of test chemicals, the *in vitro-in vivo* correlations were generally low (typically less than 0.6) and the 95% CIs were generally wide (often greater than ±40 MMAS units). Analyses were also carried out for six (overlapping) subsets of chemicals (30 water-soluble chemicals, 18 water-insoluble chemicals, and 12 surfactants; and 20 solids, 14 solutions made from solids, and 26 liquids). The results for one of these subsets, the surfactants, were more encouraging in that the correlation coefficients were generally higher (greater than 0.8 for some endpoints) and the 95% CIs tended to be narrower.

Several factors could account for the low precision of the predictions observed in this study: a) the choice of test chemicals; b) the protocols which were used; c) the variability in the *in vivo* data; d) the use of the MMAS as the *in vivo* endpoint; and e) the choice of statistical methods (correlation analysis and simple linear regression). It is now recognised that the variability of the *in vivo* data is particularly important - computer simulations carried out by Bruner *et al.* (1996) have shown that even if the alternative methods were perfectly reproducible (if their coefficients of variation were 0), the variability in the Draize scores alone would restrict the Pearson correlation coefficients to the range from 0.89-0.95 when the Draize scores are between 0 and 100, and to the range from 0.65-0.80 when the Draize scores are between 0 and 40 (typical of cosmetics ingredients).

In summary, none of the nine tests was sufficiently predictive of *in vivo* eye irritancy for the full set of test chemicals, even though the tests were sufficiently reproducible. In spite of the disappointing results, the EC/HO study made a valuable contribution to the validation process by highlighting the importance of optimising the protocols and refining the prediction models (PMs) of alternative methods before entering them into a large-scale validation study. The optimisation of protocols and refinement of PMs is now carried out routinely as part of the prevalidation process (Curren *et al.*, 1995).

The COLIPA study

The European Cosmetic, Toiletry and Perfumery Association (COLIPA) established a validation study to determine whether currently available *in vitro* methods are valid for predicting the eye irritation potential of cosmetic ingredient and formulations (Brantom *et al.*, 1997). Specifically, the study was designed to determine whether the data obtained by alternative methods could: a) provide acceptable agreement with the MMAS; b) provide acceptable agreement with individual tissue scores and recovery time in the Draize test; or c) predict correctly eye irritation potential in the rabbit eye. The COLIPA validation study was designed to build upon lessons learned in the EC/HO study, for example, by ensuring that PMs were defined before the validation study began.

Ten alternative methods were assessed in the COLIPA study: the chorioallantoic membrane vascular assay (CAMVA), EYTEX, the FL test, HET-CAM, the NRU assay, the

pollen tube growth (PTG) assay, the neutral red release test (Predi-Safe™), the RBC assay, the SM assay, and the tissue equivalent assay (TEA). Five of these tests had common protocols with the EC/HO study (the EYTEX, HET-CAM, NRU, RBC and SM tests).

The alternative methods were evaluated under blind conditions by using 55 test substances, of which 23 were cosmetic ingredients and 32 were formulations. Twenty of the cosmetic ingredients were common to the EC/HO study, so that the data from both studies could be pooled and analysed in greater detail in the future. The formulations included make-up products, skin cleansers, sunscreens, hair dyes, shampoos, deodorants and toothpastes.

The COLIPA study was carried out in two stages: a dry run on 10 test substances to ensure compliance with the standard operating procedures, and a main run on all test substances. Good laboratory practice (GLP) was used at all stages of sample coding, randomisation and supply to the participating laboratories. The raw data were collected centrally and received a quality assurance check before they were independently analysed, using statistical methods which had been agreed before the start of the study.

Using predefined criteria of reliability and relevance, the results indicated that none of the methods entered into the study could be confirmed as a valid replacement for the Draize eye test across the full range of irritancy. However, three methods (the FL test, the RBC assay and the TEA) each satisfied one criterion of reliability or relevance. The FL test and the TEA were conducted in only two laboratories, so it was concluded that their reproducibility should be checked in a further study. The predictivity of each method was assessed against the PM derived by the lead laboratory for that method. The PM for the TEA was a mathematical equation for the prediction of MMAS, associated with a 95% CI, whereas the PMs for the RBC assay and the FL test were classification models (predicting three levels of irritancy), associated with a 95% CI of the kappa (k) statistic (the k statistic is a chance-corrected measure of agreement ranging from zero [no agreement] to one [perfect agreement]). The TEA PM was capable of predicting the MMAS of moderate and severely irritating substances, but was less predictive for non-irritating and slightly irritating substances. It was therefore concluded that further analyses should be carried out to determine whether the TEA is capable of distinguishing between substances of low irritation potential. The FL test PM was adequate for distinguishing non-irritating substances from strongly irritating substances, but since substances of moderate irritancy were under-represented, it was concluded that further work should be conducted to determine whether the FL test is capable of distinguishing between substances of moderate irritancy and substances at the extremes of the irritancy scale.

In summary, the COLIPA validation study was designed to build on the results obtained in the EC/HO study, taking into account the lessons already learned. The outcome was promising for three methods (the FL test, the TEA and the RBC assay), but firm conclusions regarding their validity could not be made, indicating the need for additional, more focused, analyses.

The BGA/BMBF study

During 1988-1992, a validation study was carried out in Germany (Spielmann *et al.*, 1993 & 1996) to evaluate the suitability of two *in vitro* tests to replace the Draize eye test for severe eye irritants - the HET-CAM test and the NRU test using 3T3 (mouse fibroblast) cells (3T3 NRU). These tests were chosen for validation because they had been identified as the most promising tests for identifying severe eye irritants in an earlier project (Künstler *et al.*, 1987; Spielmann *et al.*, 1991). The validation study was coordinated by the Centre for Documentation and Evaluation of Alternative Methods to Animal Experiments (ZEBET) at the

Bundesgesundhütsamt (BGA), and was supported financially by the German Department of Research and Technology (BMDF). The study was conducted in two phases: phase I (1988-1990) consisted of a prevalidation study and a blind trial, and phase II (1990-1994) consisted of a database development phase and biometrical analysis.

During phase I, standardised protocols for 3T3 NRU and HET-CAM tests were developed, and the two tests were established in 13 laboratories. Following an independent assessment of the intralaboratory and interlaboratory reproducibilities of the two tests, 34 test chemicals were selected for the blind trial. These chemicals were supported by high quality *in vivo* data, and included chemicals outside of the limited group of surfactants for which the tests had been developed. The 34 chemicals were coded, and the two tests were assessed under blind conditions in 13 laboratories. Both tests had satisfactory intralaboratory and interlaboratory reproducibilities, although the 3T3 NRU test was more reproducible than the HET-CAM test. In contrast, the HET-CAM test was better at identifying severe eye irritants (chemicals classified as R41 according to EU guidelines). Both tests were capable of ranking the test chemicals in a similar order to that derived from *in vivo* data.

During phase II, further evaluations of the HET-CAM and 3T3 NRU tests were conducted by testing each method with 166 industrial chemicals under blind conditions in two laboratories. These chemicals, chosen to be representative of the chemicals produced by the pharmaceutical and chemical industries, were different to the 34 chemicals tested in Phase I. Thus, the HET-CAM and 3T3 NRU tests were evaluated using a total of 200 chemicals (147 new chemicals and 53 existing chemicals). During an independent quality control of the database, 57 chemicals were excluded from further analysis because of the unacceptable quality of their *in vitro* or *in vivo* data, leaving 143 chemicals for analysis. The PMs which had been developed in Phase I for the two *in vitro* tests were found to be insufficiently predictive of severe eye irritancy because a new criterion had been introduced in the EU classification system (EC, 1992) - the presence of irreversible damage within a 21-day observation period was now sufficient for an R41 classification to be assigned. Therefore, it was decided to carry out some *post hoc* data analyses using linear discriminant analysis (LDA) to obtain the optimal combination *of in vitro* endpoints for discriminating between severe (R41) and non-severe irritants. During the database development phase, a total of ten *in vitro* endpoints had been determined (nine for the HET-CAM test and one for the 3T3 NRU test), but since the full set of values for the ten endpoints was not available for 27 of the 143 chemicals, the LDA was based on 116 chemicals. This revealed that the best endpoint for identifying severe irritants was the detection time of coagulation - not only was this more predictive than the other nine endpoints, it was also more predictive than the traditional HET-CAM endpoint based on the combined use of haemorrhage, lysis and coagulation. For water-soluble chemicals, it was found that the detection time of coagulation using a 10% solution had the highest discriminating power, whereas for less water-soluble chemicals, the detection time of coagulation using the undiluted chemical was more appropriate. The classification of water-soluble chemicals was improved further by combining the time-to-coagulation endpoint with the 3T3 NRU endpoint (IC50; the concentration of test chemical resulting in a 50% inhibition of neutral red uptake). The classification models derived by LDA were confirmed by cross-validation.

The BGA/BMBF study led to the conclusion that chemicals can be classified as severe irritants (R41) with sufficient reliability by the combined use of the HET-CAM test and the 3T3 NRU test, both of which meet the requirements for well-validated tests, as defined in OECD Guideline 405. The report of validation study (Spielmann *et al.*, 1996) illustrates the

use of multivariate statistics in the development of PMs and in the design of tiered testing strategies.

The CTFA study

The Cosmetics, Toiletries and Fragrance Association (CTFA) conducted a six-year program (1990-1996) to evaluate promising *in vitro* alternatives to the Draize eye irritation test (Gettings *et al.*, 1991, 1994 & 1996). The program was carried out in three phases, each phase serving to investigate the performance of approximately 24 *in vitro* tests (not counting variations of each test) with respect to a specific group of products: in phase I, hydro-alcoholic formulations (10 materials) were tested; in phase II, oil-water emulsions (18 materials) were tested; and in phase III, surfactant-based formulations (25 materials) were investigated. All test materials were coded by an independent laboratory so that both the animal and *in vitro* tests could be conducted in a blind fashion. The animal experiments, which generally used six 6 rabbits per test material, were carried out either in parallel (phase I), or according to a randomised block design (phases II & III). The latter method enabled the *in vivo* variability of the Draize MMAS to be assessed more realistically. In all of the animal experiments, anaesthesia was applied to the eyes prior to dosing.

When the experimental stage of each phase of the study had been completed, the chemical identities were revealed, and the relationship between the *in vivo* and *in vitro* data was analysed by statistical methods. First, a "concordance analysis" was carried out in which a comparison was made between the materials which were statistically separated by their Draize scores and the materials which were statistically separated by their *in vitro* scores. The *in vitro* tests which performed to a certain level in the concordance analysis were subsequently analysed by non-linear regression to approximate the relationship between the *in vitro* and *in vivo* scores. A novel feature of this analysis was the inclusion of a 95% prediction interval. This reflects the variability of both the *in vitro* test and the *in vivo* test, and enables the observer to visualise the range of *in vivo* scores predicted by a given *in vitro* result.

The variability of the Draize test in each phase of the study was quite striking, even though the animal tests had been carried out in a single laboratory. In phases I & II, the variability was smallest for the least irritating chemicals, and increased as the irritancy increased. In contrast, the variability in phase III was greatest in the middle of the irritancy range and was smallest at the two ends of the scale. The Draize scores were confined to the lower end of the Draize scale (less than 46), which is the most relevant range for cosmetic formulations.

The performance of the *in vitro* tests also varied between the three phases. In general, the concordance of the *in vitro* assays was higher, and the prediction intervals were narrower, for phase I and III materials than they were for phase II materials.

In conclusion, several features of the CTFA study are noteworthy: a) the variability of the *in vivo* scores was taken into account when determining the performance of the *in vitro* methods; b) regression analysis was used to determine 95% prediction intervals for the estimation of *in vivo* scores; and c) the predictivity of each *in vitro* method was shown to vary according to the type of material being investigated.

The IRAG study

The Interagency Regulatory Alternatives Group (IRAG), which is made up of representatives from three US regulatory agencies (the Food and Drug Administration [FDA], the Environmental Protection Agency [EPA], and the Consumer Product Safety Commission [CPSC]), carried out a three year program (1991-1994) to evaluate the performance of *in vitro* assays for eye irritation (Bradlaw *et al.*, 1997). The evaluation was based on existing animal and *in vitro* data, which were submitted in parallel by laboratories around the world. Over 60 data sets from 41 laboratories were received for 29 different test methods. The *in vitro* data were compared not only with the MMAS, but also with the individual tissue scores representing the damage of the cornea, conjunctiva and iris.

A set of guidelines was developed to standardise the data submissions and to facilitate their review (Scala & Springer, 1997). These guidelines included: a) general guidelines for the acceptance of data; b) criteria for the collection and collation of *in vitro* data; c) criteria for the collection and collation of *in vivo* data (individual animal and tissue scores were requested); d) criteria for the review and evaluation of data; and e) the format to be used when reporting the summary of an evaluation (see below).

Five working groups were established, each containing 4–10 members, to review data from: a) organotypic models (Chamberlain *et al.*, 1997); b) chorioallantoic membrane-based assays (Spielmann *et al.*, 1997); c) cell function-based assays (Botham *et al.*, 1997); d) cell cytotoxicity assays (Harbell *et al.*, 1997); and e) other assays (Curren *et al.*, 1997). In addition, a statistical subcommittee was formed to help in the plassning and analysis of the study (Feder *et al.*, 1997). At the end of the program, each working group published a summary of its evaluation, and presented its conclusions at an open forum (Workshop on Eye Irritation Testing, **PLACE ???**, USA; November 1993). Most of the reviews were based on scatterplots of paired *in vivo* and *in vitro* data and on regression analyses of the resulting relationships. The variability of both the *in vivo* test and the *in vitro* tests were taken into account, and was generally represented on the scatterplots by the inclusion of error bars. The reviews revealed differences in predictivity between test methods for the same types of chemicals, and between chemical types for the same test method; none of the tests showed a satisfactory performance across all chemical groups. In general, the ability to obtain strong *in vitro-in vivo* correlations was compromised by the variable nature of the animal test.

The IRAG study led to the following conclusions: a) none of the *in vitro* tests, and no combination of the tests, could completely replace the animal test; b) alternatives to the Draize test were currently being used by industry as screens in the risk assessment process for product development; and c) some of the *in vitro* models have the potential to reduce animal testing, provided that they have been validated and are conducted under well-defined conditions.

The MHW/JCIA study

In 1991, the Japanese Ministry of Health and Welfare (MHW) began a study to investigate the possibility of using alternatives to the Draize test for the safety assessment of cosmetic ingredients (Ohno *et al.*, 1994). A detailed review of 16 methods led to the selection of 12 methods for inclusion in an inter-laboratory validation study (Ohno et al, 1995), carried out under the auspices of the MHW and the Japanese Cosmetic Industry Association (JCIA).

The 12 methods assessed in the MHW/JCIA study were: a) the HET-CAM method; b) the HET-CAM-trypan blue staining method (CAM-TB); c) the RBC haemolysis method; d) the haemoglobin denaturation method (HD); e) the artificial skin models SKIN^{2TM} (ZK1100 model) and MATREX^{TM}; f) cytotoxicity on normal rabbit corneal cells (CornePack^{TM}); g) the