LRI Workshop: Applicability of Skin Sensitisation Testing Methods for Regulatory Purposes.

Brussels; 2-3 February 2010

Workshop Steering Committee:

Juan-Carlos Carrillo [1], Cedric Delveaux [2], Dorothea Eigler [3], Reinhard Kreiling [4], Petra Kern [5] and Mike Woolhiser [6]

1. Shell International BV, The Hague, The Netherlands/On behalf of CESIO (Workshop Chairman)
2. CESIO/CEFIC, Brussels, Belgium
3.Evonik Goldschmidt GmbH, Essen, Germany/On behalf of CES
4. Clariant Produkte (Deutschland) GmbH, Sulzbach, Germany/ On behalf of EFfCI
5. P&G, Brussels, Belgium/ On behalf of AISE
6. The Dow Chemical Company, Midland, MI, USA/On behalf of CESIO

Invited Expert Panel:

David Basketter [7], Silvia Casati [8], Ian Kimber [9], Matthias Peiser [10] and Vera Rogiers [11]

7. DABMEB Consultancy Ltd, Bedfordshire, UK
8. ECVAM, IHCP, EC-JRC, Ispra, Italy
9. University of Manchester, UK
10. Bundesinstitut fur Risikobewertung-BfR , Berlin, Germany
11. Vrije Universiteit Brussel, Belgium

A Long-range Research Initiative (LRI) workshop, organised by CEFIC and sponsored by CESIO (*Comité Européen des Agents de Surface et leurs Intermédiaires Organiques*), EFfCI (European Federation for Cosmetic Ingredients), and CES (*Centre Europeen des Silicones*) was held on 2-3 February 2010 at the Thon Hotel, Brussels. The workshop was attended by over 50 representatives from various stakeholder groups, including industry, academia and the regulatory authorities. The scope of the workshop was to facilitate the discussion on the strengths and limitations of both, the LLNA and the traditional guinea pig tests, their applicability domain and the current use of the LLNA as the "gold standard" under REACH. A further aspect of the workshop was to raise awareness about in-vitro alternatives research and to stress the importance of using an accurate and representative reference data-base when developing and validating such emerging methods anchored to traditional toxicological readouts for the identification of skin sensitizing chemicals.

with co-sponsorship of :

European Committee of Organic Surfactants and their Intermediates

Centre Européen des Silicones

The European Federation for Cosmetic Ingredients

Background

Skin sensitisation is an important toxicological endpoint that needs to be assessed for all chemicals. All tests currently accepted by the regulatory authorities are *in vivo* systems which differ in their methodology. Performing human tests (such as the repeated insult patch test; Hript) instead of animal tests for the generation of primary skin sensitisation data, or without good knowledge of a possible sensitisation potential, is considered unethical due to the risk of volunteers becoming sensitised.

Guinea pigs have historically been the recommended animal species and guidance for testing is laid down in internationally-agreed guidelines [OECD guideline No.406 (OECD 406) originally adopted in 1982 and revised in 1992]. Of the two test methods described in OECD 406, the guinea pig maximisation test (GPMT) and the Buehler test, the GPMT is generally considered to be the most sensitive and specific, and has been used for decades.

The more recently developed LLNA offers a number of important advantages over the guinea pig tests, particularly in relation to the '3R' (refinement, reduction, replacement) principle. The LLNA is based on measuring cell proliferation in the lymph node following dermal exposure to chemicals during the induction phase of sensitisation as a measure of skin sensitisation. It does not include elicitation, challenge-induced dermal hypersensitivity reactions and/or observations thereof. The LLNA has been widely used since its formal validation and its adoption as an OECD guideline in 2002 (OECD 429). It is now the first-choice method for *in vivo* testing under REACH and results from an LLNA are often regarded as the 'gold standard' reference for the development of alternatives/replacements for animal tests..

With the increased use of the LLNA, discrepancies between the results obtained with the LLNA and the guinea pig tests and/or human experience have been found for certain classes of chemicals (e.g. various surfactants and their precursors, unsaturated fatty acids) that are generally regarded as lacking a substantial sensitisation potential. More recently, LLNA results for siloxane materials have also come under suspicion.

These results raise questions as to whether the LLNA is more sensitive than the guinea pig test and/or human experience or if the LLNA may have limitations in its applicability domain. Improved characterisation of the test is thus required to enable a better understanding of possible confounding factors. This understanding is also very important in the light of the development of non-animal testing methods, as it is essential that data against which alternative/replacement methods are validated are accurate.

Aim and Goal of Workshop

The aim of the workshop was to develop a common understanding of the applicability of current *in vivo* tests for skin sensitisation hazard identification, and by sharing industry's experience with the scientific and regulatory communities to increase awareness of strengths and weaknesses in using these tests.

A further goal of the workshop was to promote the need for a more flexible testing strategy in skin sensitisation hazard identification and risk assessment of different chemical classes.

The meeting organisers hoped that discussion of LLNA experiences would develop a better understanding of underlying mechanisms and/or potential confounders, and would promote the need for a more flexible testing strategy in the reliable hazard identification of skin sensitisers of different chemical classes.

with co-sponsorship of :

European Committee of Organic Surfactants and their Intermediates

Centre Européen des Silicones

The European Federation for Cosmetic Ingredients

Skin Sensitisation Hazard Identification
[Chaired by Dr. M. Woolhiser]

Skin Sensitisation, the Local Lymph Node Assay and Alternative Strategies (Professor Kimber, University of Manchester, UK)

The presentation covered the biological basis for skin sensitisation, skin sensitisation assays and future approaches and challenges.

Allergic contact dermatitis (ACD) is a major societal and occupational health problem. Skin sensitisation is dependent on complex, multiple cellular and molecular interactions. In essence, exposure to and delivery of an allergen during the induction phase results in stimulation within draining lymph nodes of proliferative responses by allergen-reactive T lymphocytes. This in turn results in the selective clonal expansion of allergen-responsive T lymphocytes and increases the number of cells able to recognise and respond subsequently to that same inducing allergen. It is the selective clonal expansion of T lymphoctyes that provides immunological memory and the cellular basis for skin sensitisation. The expanded population of specific T lymphocytes recognises and responds if there is further exposure to the inducing chemical allergen (elicitation phase). The activation of T lymphocytes is associated with the release of cytokines and chemokines that collectively stimulate the influx of other leukocytes and initiate the cutaneous inflammatory reaction that is recognised clinically as ACD.

Certain hurdles have to be overcome before skin sensitisation is acquired. These include bioavailability, local trauma and the release of pro-inflammatory cytokines, hapten reaction with protein or metabolism, and immunogenicity. Skin sensitisation is a complicated process and all steps are needed for it to develop.

Guinea pig sensitisation assays have been around for a long time and have served toxicology well. However, there are many reasons why there is room for improvement, not least of these being animal welfare considerations (e.g. routine use of adjuvants and elicitation of an effect where the objective of animal welfare is to prevent suffering caused by e.g. adverse effects); in addition there are problems with testing coloured materials, subjective scoring and a lack of a dose-response protocol.

An increased understanding of the cellular and molecular mechanisms of skin sensitisation and a willingness to consider species other than the guinea pig provide opportunities to consider alternative approaches. The LLNA employs a different approach in which skin sensitising chemicals are identified on the basis of their ability to simulate lymphocyte proliferative responses during the induction phase of contact sensitisation. In essence the increased proliferation in skin draining lymph nodes

following treatment is measured using tritiated thymidine. The method was developed in the UK initially and then in conjunction with the USA.

Ten years after its initial development, the objective of accurately evaluating the relative potency for use in the risk assessment process became an important challenge. The quantitative associations between lymph node cell proliferative activity and skin sensitisation potential suggested that the LLNA might permit not only the identification of potential skin allergens, but also the assessment of relative sensitisation potency. The level chosen for determining a potential skin sensitising substance was a 3-fold increase in proliferation compared with concurrent vehicle-treated controls, and the concentration of chemical necessary to provoke this level of response was accordingly designated the EC3 value (Effective Concentration for a Stimulation Index of 3). Collaborative studies were conducted in partnerships between experimental laboratories performing the LLNA and experienced clinical dermatologists. Based on clinical judgement, chemicals were classified according to their relative induction potency and compared with EC3 values estimated from LLNA dose responses. A close correlation was found between clinical assessment of potency and EC3 values, as was illustrated by the presentation of examples of chemicals with their human classification (strong, moderate, weak, extremely weak and non-sensitising) and their LLNA EC3 values.

This led to proposals for classification of contact allergens according to potency, distinguishing between contact allergens on the basis of 10-fold variations in potency of the EC3 value i.e. extreme (< 0.1%), strong (> 0.1 - < 1%) moderate (> 1 - <10%) weak (>10 - <100%) and non sensitisers.

Currently the LLNA is considered to be quantitative, objective, and suitable for chemicals which could not be evaluated previously (e.g dyes) and faster than the guinea pig test; furthermore it has important animal welfare benefits. Based on the existing validation exercises it is around 90% in line with and 10% different from the guinea pig test results. As with any other test, however, the LLNA has limitations and false positive results can occur. Caution should thus be exercised in assuming that a positive result in the LLNA is incorrect when in conflict with a negative guinea pig test. It is important to distinguish between ‘inaccurate’ or real false positive test results and ones that are perhaps unpalatable.

With regard to the future, aspirations are of course to assess skin sensitisation potential without the use of animals. However in developing alternative approaches in such a complex area, major challenges are involved in hazard identification.

Areas that require investigation are:

- epidermal bioavailability (e.g. by (quantitative) structure activity relationship ((Q)SAR), dermal absorption analysis, skin modelling);
- protein reactivity with/without metabolic activation (e.g. (Q)SAR, *in vitro* peptide reactivity assays);
- dendritic cell activation (e.g. examining cytokine production by LC (Langerhans cells)/EC (epithelial cells), LC/DC (dendritic cells) activation/maturation, LC migration/mobilisation, identification of 'danger signals');
- immunogenicity (e.g. primary T lymphocyte activation assays).

All investigations should include aspects such as ensuring sensitivity, selectivity and accuracy of the assays, as well as a consideration of metabolic activation or inactivation, an appreciation of the mechanistic relevance of the assay, and chemical delivery in the *in vitro* assays. This is illustrated with reference to questions such as: what is the mechanistic relevance of chemical-induced changes in the DC phenotype and do these reflect what is happening *in vivo*; does T lymphocyte proliferation *in vitro* equate with sensitising activity or elicitation of allergic responses; does native or inducible protein reactivity necessarily reflect skin sensitisation potential; does the absence of epidermal bioavailability necessarily equate with absence of skin sensitising activity.

Providing a holistic picture of the entire skin sensitisation process requires integration of data. However, breaking down these complex molecular and cellular interactions into separate areas means that it is difficult to rebuild a holistic picture. Moreover, alternative approaches often do not provide any indication of potency. A measure of potency is key to quantitative risk assessment for sensitisers, as the potency of contact allergens can differ by up to five orders of magnitude.

In conclusion, an LLNA, which measures EC3 values in the mouse, provides a holistic model for the assessment of relative skin sensitisation potency in an integrated system.

Aspects in the Development of "In-vitro Based Alternatives for Sensitisation Testing" (Dr. M. Peiser, German Federal Institute of Risk Assessment;BfR)

Allergic contact dermatitis caused by low molecular weight chemicals, is a delayed hypersensitivity reaction resulting in a series of complex cellular and molecular events and involving T cell-mediated hypersensitivity reactions. The problem is that some of these events are also initiated by skin irritants and toxins following skin barrier disruption. With irritants and toxins, the keratinocytes detect DAMPs (damage associated

molecular patterns). These can be endogenous species that are generated following tissue injury, e.g. heat shock proteins (HSP). Toll-like receptors (TLRs) can sense DAMP (e.g. HSPs which are ligands for both TLR2 and TLR4). Both keratinocytes and LC have TLRs, including TLR2. DAMP generation and consequently immune cell migration to the lymph node has been shown to occur also for irritants following injury. SLS has been shown to induce HSP27 in human skin. For the development of *in vitro* assays the challenge is to select predictive biomarkers that discriminate between irritation and sensitisation and for the latter, to discriminate between the induction and elicitation phases of skin sensitisation.

Due to the increasing public and political concerns regarding the use of animals for the screening of new chemicals, the Colipa Skin Tolerance Task Force collaborates with and/or funds research groups to increase and apply an understanding of the events occurring during the acquisition of skin sensitisation. Knowledge gained from this research is used to support the development and evaluation of novel alternative approaches for the identification and characterisation of skin sensitising chemicals. At present one *in chemico* (direct peptide reactivity assay (DPRA)) and two *in vitro* test methods (cell based assays (MUSST and h-CLAT)) have been evaluated within Colipa inter-laboratory ring trials and accepted by the European Centre for the Validation of Alternative Methods (ECVAM) for pre-validation. Data from all three test methods will be used to support the development of testing strategy approaches for skin sensitiser potency predictions. Furthermore, projects are funded in the areas of epidermal bioavailability, skin metabolism and T-cell recognition.

Within the European Commission's 6th Framework Programme, Sens-it-iv is currently developing a number of novel testing strategies. One such aims at identifying the most *in vivo*-like epithelial cells (EC) and EC markers. Another is attempting to identify *in vitro* conditions supporting the most *in vivo*-like EC-dendritic cell (DC) interactions. A strategy to establish a T cell based assay capable of identifying contact allergens by detection of T cell proliferation and IFN-γ is also being investigated. In addition an *in vitro* assay of biomarkers on DC, such as CD86 and IL-8, is being developed. All testing strategies will employ *in vitro* methods.

BfR is concentrating on developing new predictive biomarkers. The CAATC (contact allergen activated T cell) assay is an approach using DC from skin and involves the characterisation of the sensitisation potency of chemicals via DC-induced expression of lineage specific T cell transcription factors and cytokines. It is based on the model that each allergen evokes a specific but not exclusive pattern of these markers that can be quantified. The advantage of the assay is that it implements barrier disruption and 'danger signals' as found *in vivo* studies. In the CAATC assay, T cell transcription factors (including T-bet, GATA3, RORC and FOXP3 discriminating T helper(h) 1, Th2, Th17

and regulatory T cells) will be quantified . All these cells are crucially involved in an ACD response.

Finally, certain substances have been tested in an *in vitro* loose-fit co-culture-based sensitisation assay (LCSA), using a co-culture of keratinocytes and DC. For some chemicals (e.g. dinitrochlorobenzene, phenol, salicylic acid) there appears to be agreement between results obtained in the LCSA and in the LLNA, while for others (e.g. sodium lauryl sulphate, hexylcinnamaldehyde, nickel) there is no agreement.

The question of when alternative *in vitro* tests are likely to be available and accepted by the regulators arises. With the LLNA, it took eighteen years from the date on which the LLNA was first 'conceived' until it was adopted by the OECD as TG429 (OECD 429). Following initial publication a number of steps have to be progressed before a method is finally adopted by the OECD. Formal validation of the LLNA included inter-laboratory development of the method, publication of a standard protocol, inter-laboratory validation studies, comparison with guinea pig data, comparison with human data and regulatory review. A similar approach is needed for the *in vitro* assays and validation will be required against suitable hazard data to meet the ambitious Cosmetics Directive deadline of 2013!

With regard to what is 'in the pipeline' at OECD, included is an update of TG429 which encompasses LLNA performance standards to streamline the validation of variation of the LLNA and two new TGs on modified versions of the 'traditional' LLNA with non-radioactive protocols. Actual replacement testing methods include tests to assess skin irritation/corrosion and phototoxicity. No validated *in vitro* assays are likely to be available by the deadline of 2013 covering all aspects of the complex mechanism of sensitisation.

[In the Q & A session that followed, Professor Kimber stressed that bioavailability was the most important factor in the development of sensitisation and queried whether the *in vitro* methods presented could mimic bioavailability. His concern related to predictability and he felt that further research was needed to decide which specific cells and which specific immunological responses to investigate. In reply, Dr Peiser agreed that further research was indeed needed, in particular to ascertain which cells were predictive for sensitisation; 8-10 different cells could be involved.

A question was raised as to the role for (Q)SARs, and the view expressed that much progress had been made in understanding chemical processes. Structural alerts and chemical reactivity towards nucleophiles should play a key role in the evaluation of chemicals for their skin sensitisation potential.]

The LLNA as the Prescribed Method under REACH
[Chaired by Dr. D. Basketter (DABMEB Consultancy)]

Comparative Sensitisation Data from Industry Based on the LLNA and Guinea Pig Testing

Under REACH, skin sensitisation is a component of the minimum standard information of Annex VII, required for all tonnage bands. The LLNA is the first-choice method of *in vivo* testing (after assessment of available human, animal and alternative data) [Regulation No 1907/2006]. Data generated using the LLNA led unexpectedly to positive results for certain classes of substances. As industry sectors affected by this issue, manufactures of surfactants (and their intermediates) and manufacturers of cosmetics ingredients (such as fatty acids) through their trade associations CESIO and EFfCI and the silicone industry through CES, worked together in gathering and generating comparative data to help explain these unexpected results. Dr. J-C. Carrillo (CESIO), Dr. R. Kreiling (EFfCI) and Ms. D. Eigler (CES) presented their data.

The following background points emerged from these representatives in relation to the LLNA:

- Skin sensitisation tests as specified in OECD 406 (GPMT) and 429 (LLNA) are performed for hazard identification/assessment and classification and labelling purposes;
- The LLNA (OECD 429) is the first-choice method for *in vivo* testing under REACH;
- Only in exceptional circumstances can another test be used for which 'scientific justification' should be provided;
- No agreed 'scientific justification' is currently available;
- In general, animal welfare considerations in the EU dictate that animal tests cannot be repeated for the same endpoint in hazard assessment of a substance;
- The LLNA is the first regulatory accepted test to be formally validated;
- Validation of the LLNA was based on comparative data in guinea pigs;
- The LLNA represents a skin sensitisation method based on the '3R' principles and can provide valuable and objective information on potency.

CESIO Data
Data on two projects were presented.

In the first project, ten non-ionic sugar-lipid surfactants were tested in the GPMT and in the LLNA, and eight substances were also tested in volunteers using Hript. Irritation was assessed in the LLNA by the percentage increase in ear thickness, in the GPMT by a pre-

with co-sponsorship of :

European Committee of Organic Surfactants and their Intermediates
Centre Européen des Silicones
The European Federation for Cosmetic Ingredients

test to determine non and minimally irritant doses, and in the Hript by the irritation index in the induction phase. With the exception of the C8-C10 glucoside, irritancy in the GPMT was not consistent with that observed in the LLNA; concentrations tested in the LLNA and associated with an increase of over 10% in ear thickness did not appear to cause a visible irritant response in the GPMT.

Of the five sugar lipid substances that were positive in the LLNA, four were irritant as assessed by > 10% increase in ear thickness. It is possible that these irritant substances triggered a non-specific reaction leading to cell proliferation in the lymph node. However, the fifth substance, positive in the LLNA, was not irritant, showing a clear dose response. This indicates that irritation alone cannot be considered a possible confounding effect. By contrast none of these substances was positive in the GPMT or the Hript.

The second project relates to an initial attempt to use alternative endpoints to distinguish irritancy from sensitisation. Eight surfactant-like substances were chosen to represent non-ionic and anionic surfactant compounds for comparative testing in the LLNA and the GPMT. Towards assessing the irritancy potential of the substances in the LLNA, an alternative marker B220 was incorporated in addition to measurement of ear thickness and weight; B220 has been proposed as an alternative marker to differentiate sensitisers from irritants although this approach has not yet been sufficiently validated. Based on the Stimulation Index (SI) from thymidine incorporation, discordant results between the GPMT and the LLNA were found with five of the seven materials tested (and one equivocal result). The alternative endpoints assessed did not explain the suspected role of irritancy in obtaining discordant results. For some of the alternative endpoints, there was no clear dose response in every case and, for example, the increase in the B220 marker did not always correlate with the SI in the LLNA.

Although the LLNA appears to be suitable for many classes of substances, data collected and generated on surfactants suggest a low level of accuracy for such compounds when compared to results in guinea pig or human experience. The data submitted to ICCVAM (Interagency Coordinating Committee on the Validation of Alternative Methods) for the validation of the LLNA suggested an accuracy of 88% based on the chemicals with which it had been evaluated. In relation to surfactants, accuracy was around 47% relative to guinea pig data and previous experience. This calls into question whether the LLNA in its present form should be the first-choice method for this class of substances when compared to results in guinea pigs or on human experience.

In conclusion, there is a clear need for a better understanding of the underlying mechanisms of 'false positives' and 'false negatives' in the LLNA and for the further development and refinement of alternative endpoints. It is essential to establish a sound data base if reliable *in vitro* tests are to be developed.

with co-sponsorship of :

European Committee of Organic Surfactants and their Intermediates

Centre Européen des Silicones

The European Federation for Cosmetic Ingredients

EFfCI Data
The first indications of unexplained findings in the LLNA with fatty acid-derived surfactant raw materials surfaced in 2004. Virtually no allergic responses were seen in the GPMT but inexplicably high SI in the LLNA indicated skin sensitisation potential. Experiments were begun in 2005 to clarify whether the unexpected results in the LLNA might reflect 'false positive' findings in terms of sensitisation.

Initial data from an LLNA/ GPMT comparative screening study using a fatty acid-diglycerol ester and its reactants, indicated that the fatty acid was the potential driver of the overall positive response in the LLNA. Further data, from testing saturated versus unsaturated fatty acids, indicated that the carbon-carbon double bonds might have an influence in the response. Differences in the degree of hydrogenation of the same basic fatty acid influenced LLNA responses and led to a significant decrease in the SI value when the saturation degree of the fatty acid was increased.

EFfCI subsequently initiated a project based on the above findings. The aim was to extend the scientific basis of possible limitations of the LLNA for classes of substances that had not been adequately represented in previous validation trials, and in particular to address the hypothesis that chemicals with carbon-carbon double bonds might result in a higher number of 'false positives' in the LLNA than in the GPMT.

Nine cosmetics raw materials, mainly unsaturated aliphatic and fatty acids, were compared in the LLNA and the GPMT. All substances were either endogenous physiological components of the human body and/or natural constituents of food and/or cosmetics, and all had a long history of widespread and safe consumer use without exhibiting any skin sensitisation potential.

Six of the test substances (oleic, linoleic, linolenic and maleic acid, squalene and octinol) were positive in the LLNA but negative in the GPMT. Fumaric and succinic acid were negative, while undecylenic acid, was positive in both; together with available human data this indicated a weak skin sensitisation potential for undecylenic acid.

The biochemical properties of the test compounds which might trigger cell proliferation were examined and it was found, *inter alia* that oleic, linoleic and linolenic acid were involved in diverse cell proliferative and inflammatory processes.

Points that emerge from the literature include the fact that all identified biochemical processes apply only to *cis*- and not to *trans*-unsaturated or saturated fatty acids. Succinic and fumaric acid, but not maleic acid, are intermediates in the citric acid cycle; of these only maleic acid is positive in the LLNA. There is no significant ear swelling response or

clinical signs of local irritation on the ears in the LLNA. Dermal irritancy alone does not explain the high SI value observed with maleic acid, since the geometric isomer fumaric and the corresponding saturated succinic acids exhibit comparable (or even higher) irritant effects in pre-tests.

With regard to classification and labelling, R43 [*May cause sensitisation by skin contact*] would be required for oleic, linoleic, linolenic, and maleic acids, squalene and octinol if the LLNA results are used for classification purposes; there would be no labelling requirements for any of these substances if the GPMT is used. No labelling would be required for fumaric or succinic acid under either method while undecylenic acid would attract R43 labelling in relation to both test methods.

The compounds tested are considered to have low or no sensitisation properties based on human experience. However further data are needed to ascertain whether the absence of reported human cases of allergic contact dermatitis, despite widespread exposure, indicates absence of hazard or absence of risk.

EFfCI concludes that the substances that caused a positive result in the LLNA, but a negative result in the GPMT, possess skin sensitisation properties to such a limited degree - if at all - that formal classification and labelling is inappropriate. Furthermore basing the classification and labelling for these substances solely on the LLNA is scientifically unjustified.

The LLNA has advantages over the GPMT in terms of animal welfare, time, and objectivity of read-out and quantification of results. However for unsaturated compounds structurally similar to the substances investigated, the suitability of the LLNA for skin sensitisation testing should be carefully considered, as the GPMT might more accurately reflect relevance to humans for this group of substances.

EFfCI therefore proposes that, in the light of the uncertainties identified, the GPMT should remain a recognised and accepted test method for the identification of skin sensitisation hazard.

CES Data
A number of product applications result in skin contact with silicones. Understanding and correctly identifying potential skin sensitisation hazard are key to CES commitment to the safe use of these products in the workplace and by consumers.

A surprising number of ‘positive’ results were generated with some prototype silicones when the LLNA was validated in-house. Since such substances have thus far not been linked to skin sensitisation, CES felt that clarification of these findings was required.

The skin sensitising potential of five silicones was investigated using the GPMT and the LLNA. All five were negative in the GPMT. In the LLNA, one was negative, one positive (but strongly influenced by irritancy) and three were weakly positive, with SI values above 3. Based on the current guideline and considering only the LLNA data, classification and labelling as skin sensitisers would have been applicable to four out of five substances. Considering the GPMT data alone, none of these would require classification and labelling.

In the Weight of Evidence (WoE) analysis that was subsequently carried out, factors such as absence of occupational allergic contact dermatitis in workers with daily skin exposures to silicones over more than 10 years, the low dermal penetration potential of these silicones, their physico-chemical properties, and the absence of significant functional groups in the structures known to be associated with skin sensitisation response, were considered. Furthermore, the excessive level of irritancy in the one clearly positive result suggests a 'false positive', particularly as there is an absence of skin sensitisation response in the GPMT at high induction/challenge concentrations. Weight of Evidence thus suggests that none of the silicone materials tested represents a skin sensitisation risk to humans under normal conditions of use.

The LLNA produces questionable results for this class of materials and is thus not considered as a first-choice method for skin sensitisation testing. The future testing strategy for other silicone materials will be considered on a case-by-case basis.

The results appear to suggest that weak or non-irritant substances can lead to non-specific cell activation, resulting in a positive outcome in the LLNA. This needs to be studied further, particularly in relation to the use of the LLNA with a new substance, where little information will be available to enable a WoE approach to be used in its evaluation. Finally there is the question of how much data are considered sufficient to be referred to as 'scientific justification'.

C &L Framework and REACH and Skin Sensitising Potential

N. Ball (representing CESIO) addressed the practical implications to industry in relation to skin sensitisation potential of requirements under the 'classification and labelling' framework of REACH. The challenges involved in deriving the 'correct' skin sensitisation classification for humans were outlined.

Current European legislation requires chemical substances and preparations to be classified and labelled according to their intrinsic hazard, thus allowing appropriate risk management. Correct classification is therefore most important to enable soundly-based

advice to be given to workers and consumers. Precautionary classification is not ideal and is inconsistent with the WoE approach that is encouraged in assessing available data.

With regard to REACH and the generation of new data, the LLNA is the first-choice method for *in vivo* testing. Other approved tests (GPMT/Buehler) can be used only under 'exceptional circumstances' and where 'scientific justification' supports the use of a test other than the LLNA. However, currently there is no guidance on what justification would support the use of a method other than the LLNA. Existing human data are considered acceptable, but no new studies are allowed to be carried out.

In using a WoE approach, both the LLNA and the guinea pig assays are considered acceptable. The LLNA has been formally validated by comparing it with existing guinea pig and human data where it was found to be as predictive as the guinea pig assays; this implies that the guinea pig assays have also been validated, but no formal validation has been carried out. The tests do not assess the same endpoint; LLNA assesses induction of skin sensitisation, while the guinea pig and human patch tests assess elicitation.

Various factors influence a WoE interpretation. Available data for new substances would be predominantly from the LLNA, but with certain chemistries (e.g. surfactants, fatty acids, silicones) the LLNA and the guinea pig data are often contradictory. A question arises as to which assay is of most relevance to humans. For example, validation of the LLNA identified sodium lauryl sulphate (SLS) as a 'false positive'. However, if SLS was a 'new substance' it would only be tested in the LLNA and would thus be classified, and subsequently labelled, as a sensitiser.

The guinea pig assays also have deficiencies, as indeed do human data. Concentrations employed in the guinea pig assays are often lower than those in the LLNA, there is no threshold, many of the data are 'old' and, finally, there is a degree of subjectivity in assessing the skin responses. Human data varies widely, from well-conducted patch tests to general experience from the handling of products by consumers and industrial workers. Human patch tests often involve low doses and small groups: reported data from human experience varies widely.

An appropriate classification based on WoE needs more than one data point; this is simply not available with many chemicals. Often, the only option is to rely on the results of the assay that shows a positive result, even if it might be a 'false positive'; in many cases there are insufficient data to make a robust interpretation. Data on protein reactivity, dermal penetration and structural alerts would complement a WoE approach.

In conclusion, to derive the skin sensitisation classification most relevant to humans, additional insight is needed into what causes these 'false' results in the assays and to

ascertain their relevance to humans. In addition more ‘tools’ are needed to construct a WoE case. Finally guidance is needed on ‘applicability domains’, to help provide clarification as to whether or not certain methods are in fact appropriate for the substances in question.

with co-sponsorship of :

European Committee of Organic Surfactants and their Intermediates

Centre Européen des Silicones

The European Federation for Cosmetic Ingredients

Round Table Discussion

There was some discussion as to whether the testing conditions of the assays reported by the previous speakers had to some extent influenced the unexpectedly positive results described or whether some might be related to impurities or by-products.

Another potential issue was the comparison of a treatment at 100% test substance to naïve and/or vehicle-treated mice.

In relation to the fatty acids, there are structural variations between the *cis* and *trans* forms of unsaturated compounds that may lead to differences, for example, in metabolic oxidation and sensitivity to steric hindrance. However, this cannot explain the differences seen between guinea pigs and mice although there may be differences in metabolic activation in the mouse and the guinea pig skin. No reasons could be identified to suggest why the fatty acid results in the LLNA might not be regarded 'false positive' findings.

It is difficult to predict the size of data base that would be sufficient to argue that the use of the LLNA is not justified for such a class of chemicals, but it was noted that current published data, together with material in press could represent a sound test case.

With such substances, the general feeling of some participants is that the guinea pig tests are more predictive of risk to the worker or the consumer than the LLNA and should be the preferred test method for closely-related compounds. This is in contrast to REACH under which the LLNA is the first-choice method of test. Also regulations in relation to basic raw materials require labelling based on hazard, rather than risk.

Other points mentioned and debated include the following:

- Where there is no obvious reason to suspect from its chemistry that the substance is a sensitiser, it might be useful to evaluate what is happening in the lymph node, as effects may not be related to sensitisation. However, there is no clear strategy on how to achieve this beyond use of the B220 marker.
- LLNA does not routinely give 'false positive' results based solely on irritant potential. Many sensitising substances are irritant, some profoundly so, and these are correctly identified in this assay.
- LLNA relates to the potential of chemical substances to induce sensitisation while the guinea pig tests and human data relate to elicitation. The tests therefore assess

with co-sponsorship of :

European Committee of Organic Surfactants and their Intermediates

Centre Européen des Silicones

The European Federation for Cosmetic Ingredients

different endpoints. The guinea pig methods do, however, offer the opportunity to rechallenge and investigate questionable results.

- What constitutes 'adequate human data'? There is general guidance in the REACH documentation on what constitutes human data, but little advice is given on 'adequate' data.

[Day 2. Recap Session]

At the beginning of the second day of the workshop participants were reminded that the aim in particular was to develop a common understanding of the applicability of the current animal tests (OECD 406 and 429) and to promote a scientifically-sound and flexible testing strategy.

On the previous day the scientific background of the LLNA was presented. It is clear that considerable progress has been made with regard to the LLNA as compared with guinea pig tests, particularly with respect to the '3R' principle. However, there still are some areas of concern relating to its use. It is important to understand the limitations of the LLNA and to try to explain apparently inconsistent results. I*n vitro* methods are unlikely to be on the regulatory scene in the near future, since pre-validation activities at ECVAM are only just commencing.

There are a number of questions still to be resolved. These include a need to improve the definition of the 'applicability domain' of the guinea pig methods and the LLNA, to clarify how to deal with discordant results and to identify the next steps to improve LLNA methodology. Other unresolved areas include the evaluation of how/if human experience can be integrated into the assessment of sensitisation potential, the evaluation of 'false negative' or 'false positive' results in the validation of alternative methods, and the incorporation of test limitations into testing strategies.

Break-out groups

Workshop participants were allocated to break-out groups under specific chairpersons. Prior to breaking into separate groups the audience was briefed on the key topics to be discussed in their respective break-out groups. These included:

- Is there a single 'gold standard'? Chaired by Professor V. Rogiers (VUB. *Vrije Universiteit Brussel*)
- Applicability domain, what can we do about it? Chaired by Dr. D. Basketter (DABMEB Consultancy)
- The use of LLNA for the development of next generation tests. Chaired by Dr. S. Casati (ECVAM)

Report, Conclusions and Recommendations from Break-out Groups

Is there a single 'gold standard'?
Background:
The question was whether there is in fact a 'gold standard' for current methodology, including *in vitro* testing involving the non-intact immune system (skin penetration, peptide reactivity, danger signals, surface markers) as well as *in vivo* testing on the intact immune system in man (Hript) and in animals (Buehler, GPMT, LLNA).

- The sub-group discussed the criteria against which new *in vitro* tests should be validated, and what argumentation or scientific data are needed to support the use of the LLNA alone or in combination with the GPMT or using a WoE approach. Accepting that for many chemicals only data from the LLNA will be available in the future, the sub-group discussed how to define a WoE approach and considered whether data such as SAR and general information on the chemical classes, impurities or residues would be sufficient.

Conclusions and recommendations

- There was general agreement within the sub-group that a single 'gold standard' as such does not exist, either for validation of *in vitro* methods or for classification and labelling purposes. The 'gold standard' is instead a set of data providing a sufficient WoE.

- For guidance on method selection, it is envisaged that a peer-reviewed scientifically-based decision tree is developed, based on for example, purity, chemical structure, physico-chemical properties and existing knowledge, including knowledge of the 'applicability domain'.

- The main problem is in relation to new substances. Also regulatory acceptance of the decision tree argumentation is not evident.

- To progress matters, one path would be to organise an international workshop involving different sectors of industry, with the objective of analysing the experience gained with the LLNA and the GPMT, and to publish the results of the meeting.

Applicability Domain. What Can We Do About It?
Background:

On what constitutes an 'applicability domain', it was recognised that all toxicology tests have limitations. These may involve physically unsuitable materials and solubility considerations, but may also include specific types of chemistry that are documented as not having been assessed correctly. Alternatively, the test may be suitable for certain substances but has not been formally validated for that purpose. By describing such limitations 'applicability domains' are developed.

A Q(SAR) approach has been defined, for example, which can be used to predict the skin sensitising potency of aldehydes and ketones, but not to predict the activity of alcohols or esters i.e. the applicability domain of the Q(SAR) is defined by the chemical species, in this case aldehydes and ketones, used in its development. The LLNA has merits over the guinea pig tests in that a numerical value and some information on potency is obtained, but the limitations of both types of *in vivo* assays consequently limit the scope of the Q(SAR)s which can be defined.

On the other hand, the LLNA has been developed with over 200 substances as an alternative to guinea pig tests and is formally accepted as a complete replacement for such tests by ICCVAM/ECVAM. In contrast to Q(SAR)s, the challenge here is to try to discover and then fully characterise those domains where the LLNA does not apply.

Conclusions and recommendations
Views were expressed in the sub-group that the test substances used to validate the LLNA effectively represent the 'applicability domain' of the assay and that compounds whose chemical properties are not represented in the validation set are outside the 'applicability domain'.

It was considered by others that the model encompasses all substances and not just those used in its formal validation i.e. as the LLNA has been validated based on experience with more than 200 chemicals, the default for practical purposes is that the method will work across a broad range of chemistry. As no exceptions were defined at the point of validation, the test should be valid for all chemicals unless proved otherwise. Doubts were expressed that 'applicability domain' alone would be a valid argument at a regulatory level in relation to LLNA results.

There was general agreement in the sub-group that the way forward was to:

- develop an opinion paper based on one or two convincing dataset(s) to determine whether the LLNA is an appropriate assay if there is concern for a chemistry which is novel to the 'applicability domain';
- suggest a research programme towards understanding the immunobiology of 'false positives';

- develop a strategy for discordant data(involving for example SAR, peptide binding) prior to undertaking any *in vivo* work;
- develop a strategy (decision tree) for method selection with an approach similar to that described by the previous subgroup.

With regard to new substances, the group concluded that it will be difficult to establish scientific evidence for negating a positive result.

The use of the LLNA for the development of next generation tests

Background

Various steps are involved in the lengthy process from development to regulatory acceptance of alternative methods and the various considerations that are applied in the validation process. One of the crucial aspects in validation is the selection of relevant chemicals with which to judge the performance of the test. It is proposed that discussion should cover such areas as:

- is the LLNA in fact the 'gold standard' for the development/validation of alternative methods;
- what use should be made of other *in vivo* data (guinea-pig and human data);
- should potency be considered for development/validation purposes;
- interpretation of alternative test results that are not concordant with LLNA classifications.

Conclusions and recommendations

The sub-group concluded that the 'gold standard 'is in fact the best available evidence, (including LLNA, guinea pig and human data) and that this applies to all substances (including, for example, nickel and sodium lauryl sulphate). This approach is currently used by the authorities in selecting chemicals for the validation process where there is a need to select test items that are unequivocally considered skin sensitisers or non sensitisers. The LLNA, in contrast to the guinea-pig methods, provides information on the relative potency of a chemical. It is thus considered the primary reference method for the development/validation of *in vitro* tests designed to contribute to the characterisation of relative potency.

In addition the sub-group felt it was of value to propose some general guidelines to interpret results which fall outside the known 'applicability domain' of the LLNA. Towards establishing whether a finding is a true 'false positive' or 'false negative' consideration should be given to the presence of structural alerts in the molecule and to the absence of confounding factors such as:

- inappropriate concentration tested (i.e. not sufficient or too high);
- effect of contaminant or by-product;
- possible oxidation of the test chemical or metabolic activation/inactivation;

- unsatisfactory conduct of the LLNA;
- potential interaction of the substance with the test vehicle.

Only after confounding factors are excluded will it be of value to consider the possibility of using mechanistically-relevant *in vitro* tests to develop a better understanding of the mechanisms that are occurring in humans and to characterise the response in the animal model.

Open Forum Discussion

Inter alia the way forward was discussed.

- There was a proposal to organise an international conference, perhaps sponsored by EPAA (European Partnership for Alternative Approaches to Animal Testing) where all sectors of industry (e.g. chemicals, biocides, plant protection products, cosmetics, pharmaceuticals, detergents, food additives) could convene to discuss all available data generated on the LLNA, including those that 'fit' and those that do not. Such an open, scientific meeting could establish that in some cases the LLNA is not giving the 'correct' result. It is important to address this point before other alternative methods in the field are validated against the LLNA.

  There was agreement that such an international cross-sectorial meeting could indeed present a good possibility for making progress

- Another way forward was proposed, namely through an in-depth analysis of one or two examples of 'supposed' false positives in the LLNA, This would allow some of the important scientific and regulatory issues to be addressed.
  The first step in such an approach is to verify whether the result under consideration is indeed a 'false positive'. This could be done by providing, for example, clear evidence that the substance is not a skin sensitiser in humans, that the LLNA was conducted and interpreted properly, and by excluding any contribution to discordant results from skin sensitising contaminants or from oxidation and/or skin metabolism.
  If the substance is identified as being a false positive, then the next stage could be to develop a detailed understanding of the reasons why the substance behaved differently at the cellular and molecular level in mouse skin.

General Conclusions

In conclusion the workshop had presented a useful forum for various industry sectors to share experiences relating to the applicability of skin sensitisation test methods. Valuable recommendations were formulated on the way forward.

with co-sponsorship of :

European Committee of Organic Surfactants and their Intermediates

Centre Européen des Silicones

The European Federation for Cosmetic Ingredients

- On REACH-related issues, invited speakers suggest an approach for method selection based on a decision-making tree, thus providing a valuable tool based on scientific facts.

- With regard to the problem of apparent 'false positives' observed when using the LLNA, there was a strong recommendation to develop and publish a flowchart , addressing in a number of specific cases all potential confounding factors, thus bringing evidence to the regulators that certain substances, although positive in the LLNA, are not human sensitisers. In the longer-term, the aim would be to develop a supporting comprehensive 'gold standard' dataset.

- A clear need exists for industry to exchange ideas and to avoid re-inventing the wheel. The suggestion of organising an international inter-industries workshop to gather all available data represents a huge but possibly feasible challenge. Stakeholders may be willing to work towards such an objective as has been seen previously (e.g. meeting on applicability domain of the embryonic stem cell tests).

- A high level overview on the development of *in vitro* tests was presented. However, it became clear that there is still a long way to go before such testing is established, validated and included in international guidelines.

Finally thanks were expressed to all speakers and participants for their constructive contributions to the workshop.

References for further reading [1-22] (not an exhaustive list):

1. Aeby, P., et al., *Identifying and characterizing chemical skin sensitizers without animal testing: Colipa's research and method development program.* Toxicol In Vitro, 2010. **24**(6): p. 1465-73.
2. Basketter, D.A., G.F. Gerberick, and I. Kimber, *Strategies for identifying false positive responses in predictive skin sensitization tests.* Food Chem Toxicol, 1998. **36**(4): p. 327-33.
3. Basketter, D.A. and I. Kimber, *Updating the skin sensitization in vitro data assessment paradigm in 2009.* J Appl Toxicol, 2009.
4. Basketter, D.A., et al., *Nothing is perfect, not even the local lymph node assay: a commentary and the implications for REACH.* Contact Dermatitis, 2009. **60**(2): p. 65-9.
5. Basketter, D., et al., *Application of a weight of evidence approach to assessing discordant sensitisation datasets: implications for REACH.* Regul Toxicol Pharmacol, 2009. **55**(1): p. 90-6.
6. Basketter, D.A. and I. Kimber, *Skin sensitization, false positives and false negatives: experience with guinea pig assays.* J Appl Toxicol, 2010. **30**(5): p. 381-6.
7. Betts, C.J., et al., *B220 analysis with the local lymph node assay: proposal for a more flexible prediction model.* J Appl Toxicol, 2007. **27**(5): p. 506-10.
8. J.C. Carrillo, et al., *A CESIO study: Investigating the possible overestimation of sensitizing potentials using the LLNA.* in *7th World Surfactants Congress 2008*. 2008. Paris.
9. ECETOC, *Skin sensitisation testing for the purpose of hazard identification and risk assessment. Monograph No. 29*. 2000: Brussels.
10. ECETOC, *Contact sensitisation:classification according to potency. Technical Report No. 87.* . 2003: Brussels.
11. ECETOC, *Potency values from the local lymph node assay: Application to classification, labelling and risk assessment. Document No. 46*. 2008: Brussels.
12. Edwards, D.A., et al., *Screening petrochemicals for contact hypersensitivity potential: a comparison of the murine local lymph node assay with guinea pig and human test data.* Fundam Appl Toxicol, 1994. **23**(2): p. 179-87.
13. Gerberick, G.F., et al., *Use of a B cell marker (B220) to discriminate between allergens and irritants in the local lymph node assay.* Toxicol Sci, 2002. **68**(2): p. 420-8.
14. Garcia, C., et al., *Comparative testing for the identification of skin-sensitizing potentials of nonionic sugar lipid surfactants.* Regul Toxicol Pharmacol, 2010.
15. ICCVAM, *The murine local lymph node assay: A test method for assessing the allergic contact dermatitis potential of chemicals/compounds* in

with co-sponsorship of :

European Committee of Organic Surfactants and their Intermediates

Centre Européen des Silicones

The European Federation for Cosmetic Ingredients

http://iccvam.niehs.nih.gov/docs/immunotox_docs/llna/llnarepAppx/llnac1.htm. 1999.
16. Kirk M.K., et al., *Use of an enhanced local lymph node assay to correctly classify irritants and false positive substances*, in *SOT Congrees 2007*. 2007, MB Research Labs. www.mbresearch.com17. Kreiling, R., et al., *Comparison of the skin sensitizing potential of unsaturated compounds as assessed by the murine local lymph node assay (LLNA) and the guinea pig maximization test (GPMT).* Food Chem Toxicol, 2008. **46**(6): p. 1896-904.
18. Mehling, et al., *Discordant results between the LLNA and guinea pig tests: specificity vs. selectivity* in *SOT*. 2008. Seattle.
19. OECD, *OECD Guideline for the testing of chemicals: Skin sensitization: Local symph node assay*. 2010.
20. Roberts, D.W., et al., *Mechanistic applicability domain classification of a local lymph node assay dataset for skin sensitization.* Chem Res Toxicol, 2007. **20**(7): p. 1019-30.
21. Roberts, D.W., M. York, and D.A. Basketter, *Structure-activity relationships in the murine local lymph node assay for skin sensitization: alpha,beta-diketones.* Contact Dermatitis, 1999. **41**(1): p. 14-7.
22. Weltzien, H.U., *Safe cosmetics without animal testing? Contributions of the EU Project Sens-it-iv*. J. Verb. Lebensm., 2009.