Contact Dermatitis 2009: 60: 65–69
Printed in Singapore. All rights reserved

CONTACT DERMATITIS

Review Article

Nothing is perfect, not even the local lymph node assay: a commentary and the implications for REACH

David A. Basketter[1], John F. McFadden[1], Frank Gerberick[2], Amanda Cockshott[3] and Ian Kimber[4]

[1]St John's Institute of Dermatology, St Thomas Hospital, London SE1 7EH, UK, [2]Procter & Gamble, Cincinnati, OH, USA, [3]Health and Safety Executive, Bootle, Merseyside L20 7HS, UK, and [4]Faculty of Life Sciences, University of Manchester, Manchester M13 9PT, UK

For many regulatory authorities, the local lymph node assay (LLNA) is the preferred assay for the predictive identification of skin-sensitizing chemicals. It is the initial requirement for sensitization testing within the new REACH (Registration, Evaluation, Authorization and Restriction of Chemical substances) regulations in the European Union. The primary reasons for the preferment of the LLNA are the animal welfare benefits it provides compared with traditional guinea-pig methods (refinement and reduction of animal usage) and the general performance characteristics of the assay with regard to overall reliability, accuracy, and interpretation. Moreover, a substantial published literature on the LLNA is available making it appropriate for use as a benchmark against which new approaches, including *in vitro* alternatives, can be evaluated and validated. There is, therefore, a view that the LLNA represents the 'gold standard' for skin sensitization testing. However, although this is probably correct, it is important to recognize and acknowledge that in common with all other predictive tests (whether they be validated or not), the LLNA has limitations, in addition to strengths, some of which were mentioned above. Arguably, it is the limitations (e.g., the occurrence of false positive and false negative results) of test methods that are most important to understand. With respect to the LLNA, these limitations are similar to those associated with guinea-pig skin sensitization methods. Among these are the occurrence of false positive and false negative results, susceptibility of results to changes in vehicle, and the possibility that interspecies differences may confound interpretation. In this commentary, these issues are reviewed and their impact on the utility of the LLNA for identification, classification, and potency assessment of skin sensitizers are considered. In addition, their relevance for the future development and validation of novel *in vitro* and *in silico* alternatives is explored.

Key words: false negative; false positive; LLNA; REACH; regulatory classification; skin sensitization.

Accepted for publication 8 June 2008

Toxicology evolved during the past century and built on observations of the adverse effects of chemicals on human health, and the need to predict in advance the likely hazards and risks that chemicals might pose to man and the environment. The examples are manifold, but key events have included blindness caused by use of eyebrow/eyelash colourants (1), epidemics of skin allergy (2), birth defects caused by thalidomide (3), chronic diseases associated with occupational exposures, for example to vinyl chloride (4), and the appreciation that there exist associations between chemical mutagens and the causation of cancer (5). These and other events triggered the generation of batteries of predictive tests, the large majority using mammalian species (6, 7). The tests also evolved as experience was gained. Furthermore, they continue to develop and be developed and refined. Assays to predict skin sensitization potential have changed substantially in recent years (8–10), while methods to identify and characterize chemical respiratory allergens are best described as being still under development (11, 12). Many of these predictive tests will change again during the next decade to meet the requirements in the European Union to develop, where

possible, non-animal alternatives (13). There is one issue that all the tests have in common, be they *in vivo*, *in vitro*, or *in silico* – none of them is perfect.

It is important to recognize that the concepts of hazard identification, hazard characterization, and risk assessment are commonly confused. However, from even a cursory examination of the literature, it is evident that neither previously used nor currently available assays always make accurate predictions regarding the intrinsic hazardous properties of chemicals and the likelihood, where exposure is sufficient, that hazard may present a risk to human health. The purpose of this commentary is to expand on this subject, using as a specific example the first formally validated toxicology test, the local lymph node assay (LLNA) where these limitations appear to be similar to those associated with guinea-pig skin sensitization methods. Among these are the occurrence of false positive and false negative results, susceptibility of the results to changes in vehicle, and the possibility that interspecies differences may confound interpretation. In this study, these issues are reviewed and their impact on the utility of the LLNA for identification, classification, and potency assessment of skin sensitizers are considered. In addition, their relevance for the future development and validation of novel *in vitro* and *in silico* alternatives is explored.

Toxicity Assays: Sensitivity and Specificity

As an appreciation develops that chemicals may be presenting a particular type of risk to human health, be it an acute or chronic effect, the first task for toxicologists is to develop methods for detecting chemicals that possess this property. Thus, the key question to be answered is whether a test that is sufficiently sensitive can be developed. This question has various complexities, including whether the test system can perform in an acceptable way and whether it can be achieved reliably at a cost (in both animal laboratory and human resources) that permits, in practical terms, the evaluation of large numbers of test substances. For acute effects that may have an effect on all humans (for example skin irritation), the development of a sensitive procedure might be relatively simple. However, for more chronic effects, and especially those that only affect susceptible subpopulations, development of reliable approaches is very challenging. Examples of this are found in many areas of toxicology, but detailed discussion is not possible here and two examples will suffice here.

John Draize developed the rabbit skin irritation test in the 1940s to provide a way to assess the risk to human health that might result from spillage of chemicals on to human skin, notably during occupational exposure (14). The albino rabbit was selected, and 4-hr semi-occluded exposure with neat chemical was chosen to reflect the type of exposure that might occur during half a working day. Irritation being an acute effect, reactions were monitored over 3 days; severe effects (corrosion) could be observed more rapidly. Thus, this simple test in a few rabbits served to identify chemicals that could cause immediate burns or primary skin irritation. The evidence of early years (not published as far as these authors are aware) presumably demonstrated that the assay was of value in protecting human health. It is, however, very widely applied around the world in the context not only of occupational exposure but also for consumer protection (15, 16). It is for this latter group that the question of sensitivity most clearly arises. As irritation is not a binary phenomenon, an administrative (pragmatic) threshold had to be applied, which separated chemicals that were felt to be sufficiently irritating to the skin to be of concern from those chemicals that were not (17). Consumers are rarely exposed to neat chemicals, so it became reasonable to question whether the Draize rabbit test might be overly sensitive (18, 19). This is clearly the case with chemicals that are significantly irritant in the rabbit having little or no effect on human skin under identical exposure conditions (19–22). There is a similar need to apply an arbitrary threshold to data that reflect a biological continuum in other areas of toxicity testing. A good example of this is the Ames test where a threshold for positivity is based on the frequency of chemically induced mutations measured as the number of revertants compared with the concurrent control (23).

Toxicology assays are developed initially to ensure sensitivity of the detection of the important chemicals that are associated with the end-point of interest, but this is followed quickly by the key question of specificity. That is, whether the assay can discriminate effectively between positives and chemicals that should be regarded as negative. Often a balance has to be struck between sensitivity and specificity (24). As experience with an assay accumulates, new data may challenge that balance either by showing a failure to identify important causes of the relevant toxic end-point (false negatives) or by the identification of chemicals as positive, which in reality do not cause the effect (or do not cause it to a degree that would normally trigger concern). Positive results in a toxicology assay normally arise as a consequence of the activation of the intended mechanism (T-cell proliferation in lymph nodes for the LLNA;

genetic mutations in the Ames test for example), but may also arise through other mechanisms, some of which may be poorly understood. This latter group may be small in number but clearly represent false positives, although there are no doubt also false positives that operate through activation of the intended mechanism. Furthermore, the biological response whose activation is being studied may be a continuum driven by more than one mechanism and may include noise. As a consequence of all this, low-level responses and background noise necessitate establishment of a threshold to eliminate materials of no real concern. Ultimately, for all toxicology assays, a balance has to be struck, usually based on very pragmatic considerations, between maintaining a sufficient degree of sensitivity, which does not kill too much with respect to specificity.

LLNA: Sensitivity and Specificity

Let us now consider the specific example of the LLNA. Early development of the assay not only demonstrated successful identification of significant skin sensitizers (25) but also demonstrated problems with the well-known irritant sodium lauryl sulfate (SLS) that gave false positive results (26). An extensive retrospective analysis helped to substantiate the correct balance between sensitivity and specificity (27), and independent validation confirmed that this was appropriate (28, 29). Further work with a range of irritant chemical classes demonstrated that the LLNA did not routinely suffer from false positives from this type of substance (30), and generic advice was provided to aid the discrimination between sensitizers and skin irritants (31). Subsequently, further critical analyses of this issue have been published (32, 33).

Notwithstanding the above, experience has continued to accumulate with the LLNA, and some investigators have suggested that there may be classes of chemicals for which the assay is more prone to deliver a false positive result than were the guinea-pig tests that preceded it and accordingly have proposed strategies to accommodate this (34, 35). Much of the confusion arises from the difficulty of distinguishing between chemicals that are significant skin irritants, but not sensitizing, and chemicals that are significant skin irritants and are also weakly sensitizing. Consequently, proposals that require the measurement of irritant responses, such as ear swelling, are not always helpful – proving a chemical is an irritant that does not prove it lacks skin-sensitizing activity. Indeed, there is some evidence for a positive correlation between skin irritancy and skin sensitization (36). Some strategies may incorporate a supplementary alternate measure of the immune response, such as the B220 assay (37, 38). However, because such work is rarely undertaken in routine toxicology, the section below outlines a benchmark strategy for the interpretation of challenging LLNA results.

Implications for REACH

Within REACH (Registration, Evaluation, Authorization and Restriction of Chemical substances), the approach to the evaluation of skin sensitization hazards is somewhat different to current legislation (39). The LLNA takes much greater prominence as the method of first choice, with the use of any other approaches being an exception requiring full scientific justification. Consequently, and given the concerns already expressed by some workers/authors of oversensitivity of the LLNA and a suggested tendency to generate false positives (40, 41), it is of particular importance to ensure that the assay is used and interpreted with scientific rigour and with the application of a common sense approach and consideration of the weight of evidence (32, 33).

Several examples of these issues have been discussed with illustrations (26,30–33, 40, 41). It is instructive to consider a single theoretical example, SLS. Were this substance to be registered under REACH (or indeed under the older regulations), it would have a positive LLNA result, substantiated by a clear dose response, this outcome having been obtained in multiple laboratories (26). Therefore, should SLS be classified as a skin sensitizer and carry the R43 label? The answer obviously is 'no'. However, what matters is the rationale for this decision, which is based on a weight of evidence. The positive response in the LLNA is relatively close to the threshold; there is no structural alert; the material is a well-recognized skin irritant; human skin exposures, although limited by the irritation, do not appear to be associated with the induction of sensitization. Thus, if a new chemical gives a low stimulation index (i.e., <SLS), and is as irritant as SLS, and has no structural alerts, then it is a candidate false positive. Where the substance produces a level of stimulation that is greater than SLS, but is less irritant and has a structural alert, then it is more likely to be a true skin sensitizer. There are other permutations of such reasoning, particularly where other data may come into play in a weight of evidence process (e.g. guinea-pig or human predictive test evidence). Furthermore, this type of reasoning has to be applied with care – generally non-sensitizing irritants are not positive in the LLNA, but there are no absolute rules

(30, 41). However, application of a weight of evidence approach will be necessary within REACH because the LLNA, as is the case with all other toxicology tests, is associated with some incidence of false positives. Such logic has already been presented in principle in earlier publications (30, 32) but requires repeating here because the numbers of chemicals to be evaluated (>10 000) suggests that several hundreds of substances may be incorrectly labelled as skin sensitizing if an appropriate weight of evidence decision fails to take precedence over a misplaced faith in the accuracy of the LLNA.

Conclusion

No toxicology test is perfect, and each will always represent a balance between sensitivity and specificity. To make the best judgements on any toxicity end-point, including skin sensitization, it is necessary to use all the available evidence. The LLNA can make a substantial contribution to this evidence for the presence or absence of both skin sensitization hazard and for potency estimations where the result is positive (42–44). However, in making regulatory decisions, all available evidence (structural, chemical reactivity, epidermal bioavailability, guinea-pig results, and clinical and experimental human data) should also be considered. Where appropriate, LLNA results should be viewed in the light of experience with benchmark materials, for example using SLS as a classic false positive result. A detailed review of this with several practical examples is in preparation. Ultimately, it is the recognition and appreciation of both the strengths and the weaknesses of data sources that permit a toxicologist to come to the correct decision on regulatory classification of a chemical.

References

1. McCally A W, Farmer A G, Loomis E C. Corneal ulceration following use of Lash-Lure. *JAMA* 1933: 101: 1561.
2. Wilkinson J D, Shaw S, Andersen K E et al. Monitoring levels of preservative sensitivity in Europe. A 10 year overview (1991–2000). *Contact Dermatitis* 2002: 46: 207–210.
3. Collins T F. History and evolution of reproductive and developmental toxicology guidelines. *Curr Pharm Des* 2006: 12: 256–264.
4. Wagoner J K. Toxicity of vinyl chloride and polyvinyl chloride: a critical review. *Environ Health Perspect* 1983: 52: 61–66.
5. Cimino M C. Comparative overview of current international strategies and guidelines for genetic toxicology testing for regulatory purposes. *Environ Mol Mutagen* 2006: 47: 362–390 [Erratum, Environ Mol Mutagen 2006: 47: 746].
6. Jacobson-Karm D, Keller K A. *Toxicological Testing Handbook: Principles, Applications and Data Interpretation*, 2nd edition, New York, Informa Healthcare USA, 2006.
7. Barile F A. *Principles of Toxicology Testing*. Boca Raton, CRC Press, 2008.
8. Buehler E V. Delayed contact hypersensitivity in the guinea pig. *Arch Dermatol* 1965: 91: 171–177.
9. Magnusson B, Kligman A M. *Allergic Contact Dermatitis in the Guinea Pig. Identification of Contact Allergens*. Springfield, IL, Charles C Thomas, 1970.
10. Kimber I, Basketter D A. The murine local lymph node assay; collaborative studies and new directions: A commentary. *Food Chem Toxicol* 1992: 30: 165–169.
11. Kimber I, Agius R, Basketter D A et al. Chemical respiratory allergy: opportunities for hazard identification and characterisation. *Altern Lab Anim* 2007: 35: 243–265.
12. Arts J H, Kuper C F. Animal models to test respiratory allergy of low molecular weight chemicals: a guidance. *Methods* 2007: 41: 61–71.
13. Cosmetics directive EU. Directive 2003/15/EC of the European Parliament and of the Council of 27 February 2003 amending Council Directive 76/768/EEC on the approximation of the laws of the Member States relating to cosmetic products (7th Amendment to the European Cosmetics Directive). *Official J Eur Union* 2003: L66: 26–35.
14. Draize J H, Woodward G, Calvery H O. Methods for the study of irritation and toxicity of substances applied topically to the skin and mucous membranes. *J Pharmacol Exp Ther* 1944: 82: 377–390.
15. Draize J H. Dermal toxicity. In: *Appraisal of the Safety of Chemicals in Foods, Drugs and Cosmetics*, Littleton, CO, Association of Foods and Drugs Officials of the United States, 1959: 46–59.
16. Organisation for Economic Cooperation and Development. *Acute Dermal Irritation/Corrosion, Test Guideline 404*, Paris, OECD, 1992.
17. Basketter D A, Whittle E, Chamberlain M. Identification of irritation and corrosion hazards to skin: an alternative strategy to animal testing. *Food Chem Toxicol* 1994: 32: 539–542.
18. Nixon G A, Bannan E A, Gaynor T W, Johnston D H, Griffith J H. Evaluation of modified methods for determining skin irritation. *Regul Toxicol Pharmacol* 1990: 12: 127–136.
19. Basketter D A, Chamberlain M, Griffiths H A, York M. The classification of skin irritants by human patch test. *Food Chem Toxicol* 1997: 35: 845–852.
20. Robinson M K, Perkins M A, Basketter D A. Application of a 4 hour human patch test method for comparative and investigative assessment of skin irritation. *Contact Dermatitis* 1998: 38: 194–202.
21. Robinson M K, Basketter D A. Validity and ethics of the human 4 hour patch test as an alternative method to assess acute skin irritation potential. *Contact Dermatitis* 2001: 45: 1–12.
22. Basketter D A, York M, McFadden J P, Robinson M K. Determination of skin irritation potential in the human 4-h patch test. *Contact Dermatitis* 2004: 51: 1–4.
23. Organisation for Economic Cooperation and Development. *Test Guideline 471*, Paris, OECD, 1997.
24. Worth A P, Balls M. The principles of validation and the ECVAM validation process. *Alternatives to laboratory animals*. 2002: 30 (Suppl. 2): 125–128.
25. Basketter D A, Scholes E W. Comparison of the local lymph node assay with the guinea-pig maximization test for the detection of a range of contact allergens. *Food Chem Toxicol* 1992: 30: 65–69.
26. Loveless S E, Ladics G S, Gerberick G F et al. Further evaluation of the local lymph node assay in the final phase of an international collaborative trial. *Toxicology* 1996: 108: 141–152.
27. Basketter D A, Lea L, Cooper K, Dickens A, Stocks J, Pate I, Dearman R J, Kimber I. Thresholds for classification as a skin sensitiser in the local lymph node assay: a statistical evaluation. *Food Chem Toxicol* 1999: 37: 1167–1174.

28. Balls M, Hellsten E. Statement on the validity of the local lymph node assay for skin sensitisation testing. ECVAM Joint Research Centre, European Commission, Ispra. *Altern Lab Anim* 2000: 28: 366–367.
29. Dean J H, Twerdok L E, Tice R R, Sailstad D M, Hattan D G, Stokes W S. ICCVAM evaluation of the murine local lymph node assay. II. Conclusions and recommendations of an independent scientific peer review panel. *Regul Toxicol Pharmacol* 2001: 34: 258–273.
30. Basketter D A, Gerberick G F, Kimber I. Strategies for identifying false positive responses in predictive sensitisation tests. *Food Chem Toxicol* 1998: 36: 327–333.
31. Basketter D A, McFadden J, Evans P, Andersen K E, Jowsey I. Identification and classification of skin sensitisers: identifying false positives and false negatives. *Contact Dermatitis* 2006: 55: 268–273.
32. Basketter D A, Kimber I. Information derived from sensitisation test methods: test sensitivity, false positives and false negatives. *Contact Dermatitis* 2007: 56: 1–4.
33. Basketter D A Skin sensitisation: strategies for risk assessment and risk management. *Br J Dermatol* 2008 (in press).
34. Kligman A M, Basketter D A. A critical commentary and updating of the guinea pig maximisation test. *Contact Dermatitis* 1995: 32: 129–134.
35. Frankild S, Basketter D A, Andersen K E. The value and limitations of rechallenge in the guinea pig maximisation test. *Contact Dermatitis* 1996: 35: 135–140.
36. Basketter D A, Kan-King-Yu D, Dierkes P, Jowsey I R. Does irritation potency contribute to the skin sensitisation potency of contact allergens? *J Cutan Ocul Toxicol* 2007: 26: 279–286.
37. Gerberick G F, Cruse L W, Ryan C A, Hulette B C, Chaney J G, Skinner R A, Dearman R J, Kimber I. Use of a B cell marker to discriminate between allergens and irritants in the local lymph node assay. *Toxicol Sci* 2002: 68: 420–428.
38. Betts C, Dearman R J, Kimber I, Ryan C A, Gerberick G F, Lalko J, Api A M. B220 analysis with the local lymph node assay: proposal for a more flexible prediction model. *J Appl Toxicol* 2007: 27: 506–510.
39. Commission of the European Communities. Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC. *Official J Eur Union* 2006: L396/1.
40. Vohr H -V, Jurgen A H. The local lymph node assay being too sensitive? *Arch Toxicol* 2005: 79: 721–728.
41. Kreiling R, Hollnagel H M, Hareng L et al. Comparison of the skin sensitizing potential of unsaturated compounds and assessed by the murine local lymph node assay (LLNA) and the guinea pig maximization test (GPMT). *Food Chem Toxicol* 2008: 46: 1896–1904.
42. Basketter D A, Gerberick G F, Kimber I. The local lymph node assay EC3 value: status of validation. *Contact Dermatitis* 2007: 57: 70–75.
43. Van Loveren H, Cockshott A, Gebel T et al. Skin sensitization in chemical risk assessment: report of a WHO-IPCS international workshop focusing on dose-response assessment. *Regul Toxicol Pharmacol* 2008: 50: 155–199.
44. Api A M, Basketter D A, Cadby P A et al. Dermal sensitization quantitative risk assessment (QRA) for fragrance ingredients. *Regul Toxicol Pharmacol* 2008 (in press).

Address:
Dr David Basketter
2 Normans Road
Sharnbrook
Bedfordshire MK44 1PR
UK
Tel: +44-1234782944
e-mail: david.basketter@ukonline.co.uk