

Integrated testing and assessment approaches for skin sensitization: a commentary

David W. Roberts^{a*} and Grace Y. Patlewicz^b

ABSTRACT: A Bayesian integrated testing strategy (ITS) approach, aiming to assess skin sensitization potency, has been presented, in which data from various types of *in vitro* assays are integrated and assessed in combination for their ability to predict *in vivo* skin sensitization data. Here we discuss this approach and compare it to our quantitative mechanistic modeling (QMM) approach based on physical organic chemistry. The main findings of the Bayesian study are consistent with our chemistry-based approach and our previously published assessment of the key determinants of sensitization potency, in particular the relatively high predictive value found for chemical reactivity data and the relatively low predictive value for bioavailability parameters. As it stands at present the Bayesian approach does not utilize the full range of predictive capability that is already available, and aims only to assign potency categories rather than numerical potency values *per se*. In contrast, for many chemicals the QMM approach can already provide numerical potency predictions. However, the Bayesian approach may have potential for those chemicals where a chemistry modeling approach cannot provide a complete answer (e.g. pro-electrophiles whose *in cutaneo* activation cannot currently be modeled confidently). Nonetheless, our main message is of the importance of leveraging chemistry insights and read-across approaches to the fullest extent possible. Copyright © 2013 John Wiley & Sons, Ltd.

Keywords: Chemical reactivity; Bayesian; Kinetics; Mechanism-base modeling; Hydrophobicity

Introduction

This communication was stimulated by the recent publication by Jaworska *et al.* (2013) of a Bayesian integrated testing strategy (ITS) to assess skin sensitization potency. Here we comment on the approach referring to this paper, the proof of concept work presented in Jaworska *et al.* (2011) and the parallel paper by Natsch *et al.* (2013). Jaworska *et al.* (2013) presents a probabilistic method for integrating various types of 'novel' data with the aim of developing an ITS that can replace the current *in vivo* method, the mouse Local Lymph Node Assay (LLNA), while providing a prediction of skin sensitization potency suitable for use in risk assessment (as described in Api *et al.* 2008) and regulatory decision making. The predictions derived from the ITS model are compared against LLNA EC₃ values [the effective concentration of a substance to produce a stimulation index (SI) of 3] using a training set of 124 compounds and a test set of 21 compounds. The performance of the prediction model is characterized by its ability to correctly predict hazard on a Yes/No basis and in assigning chemicals to one of four broad potency categories (non-sensitizing (NS), weak (W, EC₃ > 10%), moderate (M, 10 > EC₃ > 1%) and strong/extreme (S, EC₃ < 1%).

The Bayesian approach integrates the following types of information:

- Direct Peptide Reactivity Assay (DPRA) (% depletion from lysine-based and cysteine-based peptides)
- Bioavailability parameters [LogK_{ow}, maximum free (unbound) concentration in the mid-epidermis C_{max}, Area under the total flux curve in the epidermis after 120 h of exposure AUC₁₂₀]
- Cell-based assays including outcomes from the KeratinoSens™ assay (KEC1.5, KEC3, IC50) and the Myeloid U937 skin sensitization test (MUSST) (CD86)
- *In silico* expert system predictions (TIMES-SS)

The value of the output from each of the assays or data processing exercises is associated with a contribution to the probability of a chemical being in a particular potency class and, by integrating these, an overall probability distribution can be made.

The approach has some merit if limited information is available, it can guide the user as to what new information will be most useful to generate for a given purpose. However, as we discuss below, in its present form it does not utilize the full range of predictive capability that is already available, and perhaps because of this it could be construed as being under-ambitious in aiming to assign potency categories (NS, W, M, S) rather than potency values *per se*. With potency categories there are inevitable anomalies, e.g. if compounds A, B, C and D have EC₃ values of 0.9, 1.5, 9.5 and 11% respectively, then B and C fall into the same potency category (M) while A and D lie in different potency categories (S and W, respectively), although in reality A and B have very similar sensitizing potencies, as do C and D.

Comparison with the Mechanistic Chemistry Approach

To a large extent, the findings of Jaworska *et al.* (2013) are consistent with, and lend support to, our mechanism-based

*Correspondence to: David W. Roberts, Liverpool John Moores University, School of Pharmacy and Biomolecular Sciences, Byrom Street, Liverpool, L3 3AF, UK. E-mail: d.w.roberts@ljmu.ac.uk

^aLiverpool John Moores University, School of Pharmacy and Biomolecular Sciences, Byrom Street, Liverpool, L3 3AF, UK

^bDuPont Haskell Global Centers for Health and Environmental Sciences, 1090 Elkton Road, Newark, DE, 19711, USA

analysis of potency data and our assessment of the key determinants of sensitization potency (see Alenius *et al.*, 2008; Roberts and Aptula, 2008; Roberts and Patlewicz, 2010). Table 1 references the key events as described by the OECD Adverse Outcome Pathway (AOP) for skin sensitization (OECD, 2012), together with our assessment.

The analysis summarized in Table 1 rationalizes what has been found in practice: namely that quantitative predictive models for sensitization potency can be derived by (i) assigning chemicals to reaction mechanistic domains; (ii) determining appropriate reactivity parameters and (if necessary) hydrophobicity within a mechanistic domain; (iii) deriving regression-based quantitative mechanistic models (QMMs) and using these, or mechanism-based read-across, to estimate the potency for untested chemicals. This chemistry-based approach enables potency to be predicted for many chemicals, although not all. Some chemicals may not be easily assignable to a reaction mechanistic domain, and data on related chemicals that could be used for read-across may not be available. In addition, a significant minority of known sensitizers are not directly reactive but require *in cutaneo* activation; progress in quantitatively modeling these processes is still quite limited.

The two main findings discussed by Jaworska *et al.* (2013) are just what would be expected from the above chemistry-based approach namely that:

1. Reactivity is the strongest contributing parameter to potency, and reactivity to a large extent drives the outcomes of the cell-based assays. This is in agreement with covalent modification of protein being the molecular initiating event.
2. Penetration parameters do not contribute substantially to the assignment of potency class. As we have argued elsewhere (Roberts and Patlewicz, 2009), the concept that ability to penetrate is a key determinant of sensitization potency, although widely accepted, is not supported by the available data. The penetration parameters in the analysis appear to have some, albeit limited, influence on potency, but this is in reality the effect of hydrophobicity, as quantified by Log P, (modeling

in cutaneo partitioning rather than penetration), which for some reaction mechanistic domains is a co-determinant, together with reactivity, of potency.

We need to consider a further finding in Jaworska *et al.* (2011, 2013), that for assignment to the weak sensitizer category, penetration appears to have a more significant role, and reactivity a less significant role. Having examined the lists of compounds investigated, we see that this finding should not be considered generally applicable, being largely the result of the selection of training set chemicals, and of the limitations of the reactivity assay (DPRA) used. A high proportion of the compounds in the weak category belong to the Schiff base (SB) domain. It is established that for the SB domain, hydrophobicity is a co-determinant of potency (Roberts *et al.*, 2006). Hydrophobicity being a major determinant of penetration ability leads to the perception that penetration is a significant parameter for the weak category. As the SB reaction is reversible, the water-based DPRA fails to measure SB reactivity. This is a more likely explanation for the apparently diminished role for reactivity in the weak sensitizer category.

Another aspect of the Jaworska *et al.* (2013) approach that requires comment concerns Michael acceptors. Jaworska *et al.* (2013) argue that chemicals with the direct acting Michael acceptor alert are less sensitizing *in vivo* than would be inferred from chemical reactivity data, owing to the anti-inflammatory action (reduced danger signal relative to chemicals acting by other reaction mechanisms), which increases with chemical reactivity (Natsch *et al.*, 2011). To deal with this, manual adjustments were made for direct acting Michael acceptors (this of course implies the ability to assign chemicals to their reaction mechanistic domains) in the following manner: if a chemical is predicted S (strong sensitizer), 40% probability mass was moved from S to M (moderate); if predicted M, 30% probability mass was moved from M to W class; if predicted W, 15% probability mass was moved from W to NS; if predicted NS, there was no change in the probability distribution. For the

Table 1. Key events from the Adverse Outcome Pathway for Skin Sensitization (as taken from OECD, 2012)

Key event	What it depends on (Roberts and Aptula 2008)
Bioavailability. This term has become widely used in contact dermatitis circles, not clearly defined but widely assumed to relate to ability to penetrate. Here we define 'bioavailability' as the amount of chemical per unit area that becomes available, in the course of the LLNA timeframe (or more generally in the exposure timeframe), in the regions of the skin where protein modification leading to sensitization can occur in competition with elimination processes.	Dose/unit area; exposure time. With rare exceptions, e.g. streptozotocin as discussed in Roberts <i>et al.</i> (2007), independent of the nature of the chemical.
Covalent modification of protein – the molecular initiating event.	Chemical reaction mechanism, reactivity. For some reaction mechanisms, dependency also on hydrophobicity.
Danger signal, Langerhans cell activation.	No evidence that it depends on the nature of the chemical in any way other than the covalent modification of the protein. Consequently, does not need to be considered explicitly when modeling potency by the Quantitative Mechanistic Modeling (QMM) approach.
Immunogenicity.	No evidence that it depends on anything other than the outcome of the covalent modification of protein event.

chemicals investigated, this seems a reasonable *ad hoc* way of proceeding, but there is a more systematic and quantitative way available.

First we note that the statement 'direct acting Michael acceptor[s] are less sensitizing *in vivo* than would be inferred from chemical reactivity data' is not strictly true. Potency of Michael acceptors is correlated with chemical reactivity, as quantified by measured rate constants – from the rate constants, the EC3 values can be calculated and correlate well with experimental EC3 values as encoded in the MA QMM (Roberts and Natsch, 2009) [the influence of the anti-inflammatory action is captured in the coefficient and the constant of the QMM equation]. The apparent need for the manual adjustment only arises because the DPRA data for Michael acceptors are compared together with DPRA data for chemicals reacting by other mechanisms. A better way for handling direct acting Michael acceptors is to: (i) establish, by experiment if necessary, that they really are Michael acceptors, and they are not among the known outlier sub-classes e.g. acrylates and methacrylates (Roberts and Natsch, 2009); (ii) determine the rate constants; and (iii) apply the MA QMM (Natsch *et al.*, 2011; Roberts and Natsch, 2009) to calculate the EC3 values.

Illustrative Examples of the Chemistry Approach

To illustrate how the chemistry-based approach could be practically applied, we have considered a handful of cyclic anhydrides included in the training set. While we did wish to compare the outcomes with those of Jaworska *et al.* (2013), the actual training set BN estimates were not provided in the paper or its accompanying supplementary information.

Trimellitic anhydride (TMA) is one of the training set compounds. Its allergenic properties have been extensively studied and it is frequently used as a 'standard allergen' in biological mechanism studies (e.g. Dearman *et al.*, 1991). Its LLNA EC3 is 9.2%. Could this information help infer the likely EC3 values of other related anhydrides? First we note that carboxylic anhydrides are reactive acylating agents, and are easily hydrolyzed in aqueous solution. Consequently, 24-h peptide depletion data in the lysine peptide assay at pH 10 cannot be considered as reliable measures of reactivity for anhydrides. Disappearance of the anhydrides, partly by base hydrolysis and partly by reaction with peptide, is likely to be complete well within the 24-h period. It would be quite easy to determine rate constants for the acylation of amino groups by anhydrides by appropriately designed experimental kinetics studies. In the absence of such information, we can make the simplifying assumption that acylation reactivity of cyclic anhydrides towards amino groups does not vary greatly. Second we note that although we do not have a QMM for acylating agents *per se*, it is well established that anhydrides are hard electrophiles, and on that basis we would expect that their potency would be correlated with a combination of reactivity and hydrophobicity parameters (similar to Schiff base and $S_{\text{N}}2$ electrophiles). On this basis, it would be unsafe to consider the rather low potency of TMA to be typical of anhydrides in general, as there is a strong possibility that the low potency results from the low hydrophobicity of TMA – at skin pH 5.5, the CO_2H group will be fully ionized as CO_2^- corresponding to a logP value of -3.14 for TMA.

Bearing the above in mind, what could we predict for **Phthalic anhydride (PA)** if there were no LLNA data on this compound? PA, like TMA, is easily recognizable by its chemical structure as an acylating agent, with no other type of electrophilic reactivity apparent. Profiling PA and TMA against structural alert schemes such as those implemented within the OECD Toolbox v3.1 would lead to the same conclusion. The two structures differ only in the presence or absence of a CO_2^- group – this has only a small Hammett constant (0.09 meta, -0.05 para), indicative of only a minimal effect on reactivity of the anhydride group. Thus we can regard PA and TMA as likely to be similarly reactive. The calculated logP value for PA is 1.19, i.e. 4.33 log units higher than that of TMA. Therefore, assuming potency for these compounds to be dependent on hydrophobicity, we would expect PA to be substantially more potent than TMA. To make an estimate of the EC3 value we would need to know the value of the coefficient a for logP dependency for anhydrides. We can estimate the EC3 of PA for various a values as follows.

TMA has EC3 = 9.2%, MW 192, so $\text{pEC3} = 1.32$

Hence, based on the logP difference of 4.33 we can estimate that the pEC3 of PA will be given by $\text{pEC3 (TMA)} + 4.33a = 1.32 + 4.33a$

For PA, pEC3 values can now be calculated for different a values and converted to estimated EC3 values, multiplying the molecular weight (148) by $10^{(-\text{pEC3})}$.

logP coefficient	1	0.75	0.5	0.25
EC3%	0.0003	0.004	0.05	0.65

Thus we can predict, assuming that potency is hydrophobicity-dependent, that PA will be a strong sensitizer (EC3 <1%). Confidence in this prediction is based on confidence in the assumption of hydrophobicity dependence. The observed EC3 value for PA would correspond to a value of 0.38 for a . This is close to the value (0.41) for the logP coefficient in the Schiff base QMM. This is consistent with the argument that a reflects the coefficient relating the *in cutaneo* partitioning system (X,Y) and the octanol/water system for H-polar substrates according to the Collander relationship ($\log P_{\text{X/Y}} = a \log P_{\text{o/w}} + b$) (Collander, 1947).

A third anhydride listed in the training set is **hexahydrophthalic anhydride (HPA)**. This compound has a logP value of 0.84, thus less than that of PA, but still much larger than that of TMA. Because the anhydride ring is fused to an aliphatic system, we cannot safely assume that HPA is similarly reactive to PA and TMA. In the absence of kinetic data we cannot estimate the magnitude of any difference, but we consider it a safe assumption that the reactivity differences will be outweighed by the logP difference. Thus we would expect HPA to be somewhat less potent than PA, but substantially more potent than TMA, but we cannot predict confidently which side of the strong/moderate borderline (1%) its EC3 will lie. The observed EC3 (0.84%) is in fact very close to this borderline, within the experimental limits of the LLNA.

In order to make a more precise estimate of the EC3, for HPA, or to be able to predict EC values of other anhydrides, we would need to perform kinetic experiments to determine the relative rate constants for reactions of TMA, PA and HPA with a model primary amine nucleophile. These experiments could be done with a lysine-based peptide or with a simpler model such as butylamine. In aqueous solution, hydrolysis is likely to be a

significant competing reaction, but this can be accommodated when analysing the kinetic data.

Maleic anhydride (MA) can also act as an acylating agent, but unlike PA, HPA and TMA, it can also react as a Michael acceptor. This explains why the DPRA findings reported indicate that, unlike PA, HPA and TMA, MA gives high (100%) peptide depletion with cysteine. There is also literature evidence, with kinetics, for a fast Michael addition reaction between MA and butanethiol in the presence of triethylenediamine in xylene at 25°C (Dmuchovsky *et al.*, 1964). Both carbons of the double bond have an electronegative substituent (CO.O.CO). The anhydride group being more electronegative than an ester group, we expect MA to be more reactive, as a Michael acceptor, than any of the Michael acceptors used to establish the MA QMM (Roberts and Natsch, 2009). Consequently we would expect it to fall into the strong sensitizer category. To make a more precise prediction, our approach would require determination of the rate constant for reaction of MA, relative to other known Michael acceptors sensitizers, with a thiol model nucleophile. One option may be to measure the rate constant in the Roberts and Natsch (2009) system.

These above examples highlight how a physical organic chemistry approach using kinetic data can be exploited to provide reasonable estimates of sensitizing potency. Chemistry principles can also be applied to chemicals with no expectation of skin sensitization. Isopropanol for example, which is listed in the test set, arguably does not need to be tested in any *in vitro* system as it is not electrophilic by chemical inspection, contains no functionality indicating that it would be reactive (also evident from profiling with the OECD Toolbox) and other information from other surrogate assays such as *in vitro* chromosomal aberration (CA), Mouse lymphoma and Ames as discussed by Patlewicz *et al.* (2010) are all experimentally negative (based on available information as reported in the OECD Toolbox). Accordingly this would support a non-sensitizing outcome for this chemical. Read-across from other related alcohols could add to the weight of evidence supporting a non-sensitizing outcome.

The Way Forward

Thus for a number of chemicals, it is possible to infer both sensitization potential and potency through exploiting chemistry information rather than relying on a Bayesian approach of integrating different assay outcomes. We see potential merit in the approach presented, particularly for those chemicals where a chemistry modeling approach cannot provide a complete answer (e.g. pro-electrophiles whose *in cutaneo* activation cannot currently be modeled confidently). However we do wish to reinforce the message of the importance of leveraging chemistry insights and read-across approaches to the fullest extent possible rather than relying on a prediction model where the risk is that a number of assays are run in a checkbox fashion, a probability being computed, but no reasoning being applied to challenge the relevance or validity of the outcome derived.

A key prerequisite for prediction of potency is quantitative reactivity information, enabling new chemicals to be compared against chemicals of known potency that react by the same chemical mechanism. In contact dermatitis circles it has become routine to regard the DPRA as the 'obvious' source of reactivity data for modeling purposes. However, the DPRA was developed and has been formally validated only as an assay for determining whether or not a chemical is reactive enough to be a sensitizer: it

should be noted that DPRA data are not appropriate for quantitative modeling purposes, and indeed Jaworska *et al.* (2013) point out that kinetic data, together with chemical mechanism information would be an improvement. The output (percent depletion of peptide) of the DPRA is not, for several reasons (Roberts *et al.*, 2008), well correlated with the rate constant, which is the most suitable parameter for modeling. There is currently no single 'reactivity assay' for generating rate constants for modeling contact allergenicity. In our view there is no need to develop such an assay, and in fact it would be highly undesirable to do so. Known sensitizers cover an EC₃ range of at least 10⁵ (Basketter *et al.*, 2012) and a reactivity range at least as wide and probably wider. In terms of hydrophobicity, the range is still larger. No single experimental protocol could practically cover all these ranges. There are many ways to measure rate constants, and it is a long recognized tenet of physical organic chemistry that the choice of how best to determine the rate constant for a given chemical depends on the nature of the chemical.

Conflict of Interest

The authors did not report any conflict of interest.

References

- Alenius H, Roberts DW, Tokura Y, Lauerma A, Patlewicz G, Roberts MS. 2008. Skin, Drug and Chemical Reactions. *DDT: Dis. Mech.* **5**: e211–e220.
- Api AM, Basketter DA, Cadby PA, Cano MF, Ellis G, Gerberick GF, Griem P, McNamee PM, Ryan CA, Safford R. 2008. Dermal sensitization quantitative risk assessment (QRA) for fragrance ingredients. *Reg. Toxicol. Pharmacol.* **52**: 3–23.
- Basketter DA, Clewell H, Kimber I, Rossi A, Blaauboer B, Burrier R, Daneshian M, Eskes C, Goldberg A, Hasiwa N, Hoffmann S, Jaworska J, Knudsen TB, Landsiedel R, Leist M, Locke P, Maxwell G, McKim J, McVey EA, Ouédraogo G, Patlewicz G, Pelkonen O, Roggen E, Rovida C, Ruhdel I, Schwarz M, Schepky A, Schoeters G, Skinner N, Trentz K, Turner M, Vanparys P, Yager J, Zurlo J, Hartung T. 2012. A roadmap for the development of alternative (non-animal) methods for systemic toxicity testing - t4 report*. *ALTEX* **29**: 3–91.
- Collander, R. 1947. On lipid solubility. *Acta Physiol. Scand.* **13**: 363–381.
- Dearman RJ, Mitchell JA, Basketter DA, Kimber I. 1991. Differential ability of occupational chemical contact and respiratory allergens to cause immediate and delayed dermal hypersensitivity reactions in mice. *Int. Arch. Allergy Immunol.* **95**: 70–76.
- Dmuchovsky B, Vineyard BD, Zienty FB. 1964. The Mechanism of the Base-Catalyzed Addition of Thiols to Maleic Anhydride. *J. Am. Chem. Soc.* **86**: 2874–2877.
- Jaworska J, Harol A, Kern PS, Gerberick GF. 2011. Integrating non-animal test information into an adaptive testing strategy - skin sensitization proof of concept case. *ALTEX* **28**: 211–225.
- Jaworska J, Dancik Y, Kern P, Gerberick F, Natsch A. 2013. Bayesian integrated testing strategy to assess skin sensitization potency: from theory to practice. *J. Appl. Toxicol.* DOI: 10.1002/jat.2869
- Natsch A, Haupt T, Laue H. 2011. Relating skin sensitizing potency to chemical reactivity: Reactive Michael acceptors inhibit NF-κB signaling and are less sensitizing than SNAr- and SN2-reactive chemicals. *Chem. Res. Toxicol.* **24**: 2018–2027.
- Natsch A, Ryan CA, Foertsch L, Emter R, Jaworska J, Gerberick F, Kern P. 2013. A dataset on 145 chemicals tested in alternative assays for skin sensitization undergoing prevalidation. *J. Appl. Toxicol.* DOI: 10.1002/jat.2868
- OECD. 2012. AOP for Skin Sensitization Series on Testing and Assessment No. 168, Part 1 and Part 2. OECD: Paris.
- Patlewicz G, Mekenyan O, Dimitrova G, Kuseva C, Todorov M, Kotov S, Stoeva S, Donner EM. 2010. Can mutagenicity information be useful in an Integrated Testing Strategy (ITS) for skin sensitization? *SAR QSAR Env. Res.* **21**: 619–656.
- Roberts DW, Aptula AO, Patlewicz G. 2006. Mechanistic Applicability Domains for Non-Animal Based Prediction of Toxicological Endpoints. QSAR Analysis of the Schiff Base Applicability Domain for Skin Sensitization. *Chem. Res. Toxicol.* **19**: 1228–1233.

- Roberts DW, Aptula AO, Patlewicz G. 2007. Electrophilic Chemistry Related to Skin Sensitization. Reaction Mechanistic Applicability Domain Classification for a Published Data Set of 106 Chemicals Tested in the Mouse Local Lymph Node. *Chem. Res. Toxicol.* **20**: 44–60.
- Roberts DW, Aptula AO. 2008. Determinants of skin sensitization potential. *J. Appl. Toxicol.* **28**: 377–387.
- Roberts DW, Aptula AO, Patlewicz G, Pease C. 2008. Chemical Reactivity Indices and Mechanism-based read across for non-animal based assessment of skin sensitization potential. *J. Appl. Toxicol.* **28**: 443–454.
- Roberts DW, Natsch, A. 2009. High throughput kinetic profiling approach for covalent binding to peptides: Application to skin sensitization potency of Michael acceptor electrophiles. *Chem. Res. Toxicol.* **22**: 592–603.
- Roberts DW, Patlewicz G. 2009. Chemistry based non-animal predictive modeling for skin sensitization. In *Ecotoxicology Modeling*, Devillers J (ed.). Springer: New York; 61–83.
- Roberts DW, Patlewicz GY. 2010. Updating the Skin Sensitization in Vitro Data Assessment Paradigm in 2009 – a chemistry and QSAR perspective. *J. Appl. Toxicol.* **30**: 286–288.