Regulatory Toxicology and Pharmacology xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Regulatory Toxicology and Pharmacology

journal homepage: www.elsevier.com/locate/yrtph

Evaluating the performance of integrated approaches for hazard identification of skin sensitizing chemicals

Jochem W. van der Veen [b,a], Emiel Rorije [b], Roger Emter [c], Andreas Natsch [c], Henk van Loveren [b,a], Janine Ezendam [b,*]

[a] Department of Toxicogenomics, Maastricht University, PO Box 616, NL-6200 MD Maastricht, The Netherlands
[b] National Institute for Public Health and the Environment (RIVM), PO Box 1, NL-3720BA Bilthoven, The Netherlands
[c] Givaudan Schweiz AG, Ueberlandstrasse 138, CH-8600 Dübendorf, Switzerland

ARTICLE INFO

Article history:
Received 13 January 2014
Available online xxxx

Keywords:
Skin sensitization
Integrated testing strategy
Keratinocytes
Allergic contact dermatitis
Animal-free testing
LLNA
QSAR
DPRA
Adverse outcome pathway

ABSTRACT

The currently available animal-free methods for the detection of skin sensitizing potential of chemicals seem promising. However, no single method is able to comprehensively represent the complexity of the processes involved in skin sensitization. To ensure a mechanistic basis and cover the complexity, multiple methods should be integrated into a testing strategy, in accordance with the adverse outcome pathway that describes all key events in skin sensitization. Although current majority voting testing strategies have proven effective, the performance of individual methods is not taken into account. To that end, we designed a tiered strategy based on complementary characteristics of the included methods, and compared it to a majority voting approach. This tiered testing strategy was able to correctly identify all 41 chemicals tested. In terms of total number of experiments required, the tiered testing strategy requires less experiments compared to the majority voting approach. On the other hand, this tiered strategy is more complex due the number of different alternative methods required, and predicted costs are similar for both strategies. Both the tiered and majority voting strategies provide a mechanistic basis for skin sensitization testing, but the strategy most suitable for regulatory decision-making remains to be determined.

1. Introduction

Changes in EU legislations, such as the 7th amendment to the cosmetics directive and the REACH regulation (EC, 2006, 2008), have prompted the development of alternative methods that assess skin sensitizing potential of chemicals (Adler et al., 2011; Rovida et al., 2013; Vandebriel and van Loveren, 2010). Although several of these alternatives seem promising (Zuang et al., 2013), it is unlikely that a single method can capture the complexity of the processes involved in skin sensitization. It is therefore foreseen that multiple alternative methods need to be integrated into a testing strategy for reliable classification of chemicals. The Organization for Economic Co-operation and Development (OECD) has proposed an adverse outcome pathway (AOP) that describes the molecular initiating event and key events that lead to allergic contact dermatitis (ACD) (OECD, 2012). This AOP can guide the integration of methods that each represent a different key event in the development of ACD.

The binding of a chemical to proteins is considered the molecular initiating event of skin sensitization and is influenced by bioavailability of the chemical and the cellular metabolism that either activates or inactivates a chemical. The initiating event is the basis for methods that predict sensitizing potential of chemicals through the binding capability of chemicals, such as the ECVAM-validated direct peptide reactivity assay (DPRA) (Gerberick et al., 2004; EC-JRC, 2013a) or quantitative structure–activity relationships (QSARs). Subsequently, the epithelial cells respond to the haptens and hapten–protein complexes, mainly through the production and release of cytokines and stress markers by keratinocytes (KC). This includes interleukin IL-18, which can be used to predict the sensitizing potential of chemicals in epithelial cell lines and 3D-skin models (Gibbs et al., 2013; McKim et al., 2010; Corsini et al., 2009; Van Och et al., 2005). In addition, cellular stress caused by chemicals activates cytoprotective responses such as the Nrf2–Keap1 pathway, which serves as the basis of several reporter assays, such as the KeratinoSens assay for which the ECVAM-validation report is expected soon (EC-JRC, 2013b; Emter et al., 2010). Methods that assess multiple stress related cellular mechanisms are available, such as a gene signature that includes

* Corresponding author.
E-mail address: Janine.ezendam@rivm.nl (J. Ezendam).

http://dx.doi.org/10.1016/j.yrtph.2014.04.018
0273-2300/

biomarker genes with roles in oxidative stress, immune response and gene regulation (van der Veen et al., 2013).

The next key event outlines the activation of dendritic cells (DC), which is influenced by both the chemical and the mediators released by keratinocytes. Activation leads to up-regulation of co-stimulatory molecules, such as CD86. In the h-Clat and MUSST methods this is indicative of the skin sensitizing potential of chemicals (Nukada et al., 2011). The h-Clat has been validated by EURL ECVAM and results are expected in 2014. Upon activation, DCs migrate out of the skin towards the closest lymph node and present a peptide-hapten conjugate on the MHC-class II protein, here the T-cells can recognize the haptenated peptide and start their clonal expansion. This event is measured in the LLNA, but currently no routine *in vitro* assay is available to assess this step in the AOP (Martin et al., 2010; Adler et al., 2011; OECD, 2012) .

These animal-free classification methods can be combined in a testing strategy in which molecular initiating and key events of the AOP are captured. Recently, a majority voting approach was proposed in which results from the DPRA, KeratinoSens and MUSST methods are used to classify a chemical according to the most prevalent prediction (Natsch et al., 2013; Bauch et al., 2012). In this study, we compared this majority voting strategy to a tiered testing strategy that is based on the predictive performance of the included methods.

2. Materials and methods

2.1. Chemicals

The sensitizing and non-sensitizing chemicals used in this study are shown in Table 1. The sensitizing compounds were selected based on human evidence and to reflect various potency classes. In addition, chemicals that have proven either false-positive or false-negative in the LLNA were included (Gerberick et al., 2005). All compounds were obtained from Sigma–Aldrich (Zwijndrecht, The Netherlands), except for 2-Mercaptobenzothiazole, which was obtained from Merck (Schiphol-Rijk, The Netherlands). Chemicals were dissolved in either absolute ethanol or dimethylsulfoxide (DMSO) and then added to the cells to a final solvent concentration of 1%.

2.2. Methods reflecting protein reactivity

The protein reactivity of chemicals was assessed using the individually effective methods of *in chemico* peptide reactivity methods and *in silico* QSAR methods.

2.3. Peptide reactivity

A literature search was performed to establish for which of the chemicals peptide reactivity has been assessed *in chemico* (Natsch et al., 2013; Bauch et al., 2012; Gerberick et al., 2007). The peptide reactivity of the 4 chemicals (tBHQ, TIBP, MA and HEG) that were not described in literature was evaluated using the DPRA (Gerberick et al., 2007). In short, the chemical was added to a synthetic peptide containing either a cysteine or a lysine. The mixture was then analyzed with HPLC using UV detection. After 24 h the mixture was analyzed again, a sensitizing chemical is designated by a significant decrease in the peak related to the unmodified peptide.

2.4. QSAR predictions for skin sensitization

In order to evaluate the predictive value of skin sensitization QSARs for our chemical set, the non-commercial QSAR models of MultiCASE, CAESAR, DEREK and the OECD QSAR toolbox were applied to all chemicals to generate prediction of skin sensitization potential, these models are briefly described here.

MultiCASE (Klopman et al., 2005) generates QSAR models based on substructure fragments linked to biological activity. A MultiCASE implementation for skin sensitization from the Danish EPA (DTU, 2013) is used here. In the present study, only the positive and negative predictions within the applicability domain were taken into account. The CAESAR model (Chaudhry et al., 2010) uses atom centered fragments as descriptors in a multivariate statistical model. The model gives a prediction of active or inactive (as skin sensitizer), together with applicability domain information. Again, only predictions of active or inactive within the applicability domain are taken into account. The DEREK knowledgebase (Lhasa, 2013) is a collection of structural alerts linked to skin sensitization. The model only identifies skin sensitizers and is not meant to identify non-sensitizers. Despite this limitation, we have interpreted the absence of any structural alert as a prediction of non-sensitization. DEREK predictions of "certain", "probable" and "plausible" were considered as positive predictions of skin sensitization. Predictions of "improbable", "impossible" or "nothing to report" were interpreted as a prediction of non-sensitization. Equivocal predictions were not taken into account. In the OECD QSAR Toolbox software (OECD, 2013) an implementation of a set of protein binding reactivity alerts from Enoch et al. (2008) is present, as the "Protein binding" profile. These alerts are considered indicative of reactivity towards proteins, and subsequently any substance that has an alert in this profile is considered a skin sensitizer for this study. Absence of any of the alerts in this profile was taken as a prediction of non-sensitization.

2.5. Independent Bayesian approach to evaluate a battery of QSAR predictions

Instead of characterizing the individual predictivity of the QSAR models for our dataset, a classification of the four QSAR models combined (QSAR-battery) was generated, based on Bayesian statistics. A detailed description of the methodology, applied to skin sensitization, can be found in Rorije et al. (2013). In addition, the applicability domain information provided by the individual QSAR methods is taken into account. Furthermore, a newer version of the DEREK knowledge base and an entirely new model (CAESAR) are introduced here. The specificity and sensitivity of each model used in our Bayesian analysis, taking into account the applicability domain information in the case of MultiCASE and CAESAR, are based on the analysis of a large number of chemicals in comparison with the LLNA (Table 2).

Threshold values used to determine a reliable prediction from the battery of QSARs are applied as proposed in Rorije et al. (2013). These values are >80% or >90% probability for a positive or a negative conclusion respectively. This is based on the reliability with which the GPMT test predicts the LLNA outcome (or vice versa) in the official LLNA validation study (NICEATM-ICCVAM, 1999). If there are insufficient or conflicting results from the battery of QSAR models, these thresholds will not be reached and substances are considered equivocal.

2.6. Methods reflecting epithelial response

To assess which *in vitro* approach would be most effective in a tiered strategy, the expression of biomarker genes, activation of the Nrf2 transcription factor, and production of interleukin 18 were evaluated.

Please cite this article in press as: van der Veen, J.W., et al. Evaluating the performance of integrated approaches for hazard identification of skin sensitizing chemicals. Regul. Toxicol. Pharmacol. (2014), http://dx.doi.org/10.1016/j.yrtph.2014.04.018