Regulatory Toxicology and Pharmacology 78 (2016) 45–52

Contents lists available at ScienceDirect

Regulatory Toxicology and Pharmacology

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

A data-based exploration of the adverse outcome pathway for skin sensitization points to the necessary requirements for its prediction with alternative methods

Romualdo Benigni*, Cecilia Bossa, Olga Tcheremenskaia

Environment and Health Department, Istituto Superiore di Sanita', Viale Regina Elena 299, 00161, Rome, Italy

ARTICLE INFO

Article history:
Received 21 December 2015
Received in revised form
23 March 2016
Accepted 12 April 2016
Available online 14 April 2016

Keywords:
Skin sensitization
Alternative
Human health
Modeling
QSAR
Structure-activity
Prediction
Adverse outcome pathways

ABSTRACT

This paper presents new data-based analyses on the ability of alternative methods to predict the skin sensitization potential of chemicals. It appears that skin sensitization, as shown in humans and rodents, can be predicted with good accuracy both with *in vitro* assays and QSAR approaches. The accuracy is about the same: 85–90%. Given that every biological measure has inherent uncertainty, this performance is quite remarkable. Overall, there is a good correlation between human data and experimental *in vivo* systems, except for sensitizers of intermediate potency. This uncertainty/variability is probably the reason why alternative methods are quite efficient in predicting both strong and non-sensitizers, but not the intermediate potency sensitizers. A detailed analysis of the predictivity of the individual approaches shows that the biological *in vitro* assays have limited added value in respect to the *in chemico*/QSAR ones, and suggests that the primary interaction with proteins is the rate-limiting step of the entire process. This confirms evidence from other fields (e.g., carcinogenicity, QSAR) indicating that successful predictive models are based on the parameterization of a few mechanistic features/events, whereas the consideration of all events supposedly involved in a toxicity pathway contributes to increase the uncertainty of the predictions.

1. Introduction

Contact allergy to chemicals is a recognized health problem for workers, professional and consumers. Once sensitized, an individual remains so for life, and it is therefore important to know whether or not a chemical possesses skin sensitization potential before skin contact is made. Mechanistically, it is a delayed-type of hypersensitivity reaction triggered by small reactive chemicals that can result from the induction of skin sensitization. This happen in two phases, induction and elicitation. The induction phase starts with the penetration of a chemical into the outer epidermis layer of the skin (after biotransformation, if necessary). Subsequently, the chemical or its metabolites covalently interact with skin proteins (haptenation). The haptenized chemical then triggers an inflammatory response in keratinocytes, as evidenced by the expression of pro-inflammatory interleukin 1 beta. Similar actions are observed in dendritic cells, such as Langerhans cells, which become activated. The activated dendritic cells migrate to local lymph nodes, where they present parts of the haptenized chemical to naive T lymphocytes. This results in the differentiation and proliferation of chemical-specific memory T lymphocytes. The elicitation phase is initiated upon a next contact with the chemical. Following skin penetration and haptenation, the modified chemical is taken up into dendritic cells. This recruits the pre-existing circulating chemical-specific memory T lymphocytes to the epidermis, which secrete pro-inflammatory cytokines and mobilize cytotoxic T lymphocytes. Collectively, this leads to the typical clinical symptoms associated with allergic contact dermatitis (Goebel et al., 2012; Vinken, 2013; Urbisch et al., 2015).

Within the regulatory context, traditionally the skin sensitization ability of chemicals has been tested with two established animal assays: the Local Lymph Node Assay (LLNA) and the Guinea Pig maximization Test (GPMT) (Goebel et al., 2012). In the current practice, LLNA has become the method of choice: it yields quantitative results and is suitable for classifying the sensitizers into the sub-categories 1A (strong/extreme skin sensitizers) and 1B (moderate skin sensitizers) according to the Globally Harmonized System (GHS) for classification and labeling of chemicals (United

* Corresponding author.
E-mail address: Romualdo.benigni@gmail.com (R. Benigni).

http://dx.doi.org/10.1016/j.yrtph.2016.04.003
0273-2300/

Nations, 2011).

However, recent legislation has either encouraged the reduction of animal experimentation (European substances legislation Nr. 1907/2006 [Registration, Evaluation, Authorization and Restriction of Chemicals (REACh)], or banned it completely for certain types of products (7th amendment of the Cosmetic Directive (Council Directive 76/768/EEC of 1976-07-27; now Cosmetics Regulation: REGULATION (EC) No. 1223/2009)). As a consequence, there has been intense scientific work to develop alternative methods for identifying skin sensitizers. Since the chemical and biological pathways involved are relatively well characterized, the general idea underlying such research is to mimic the main events of the hypothesized toxicological mechanism with *in vitro* assays. The key events involved have been formally described by the Organization for Economic Co-operation and Development (OECD) in a document entitled "The Adverse Outcome Pathway (AOP) for Skin Sensitization Initiated by Covalent Binding to Proteins", with the goal of facilitating the development of methods and approaches addressing the relevant events (OECD, 2012). Briefly, the AOP framework describes a Molecular Initiating Event (MIE) that includes the covalent interaction of the chemical with epidermal proteins -following skin penetration-, and 3 key events at the cellular and organ level, namely the induction of an inflammatory response in keratinocytes, the activation of dendritic cells and the proliferation of T lymphocytes. It should be noted that the animal LLNA test represents the Key Event 4 -at organ level (Lymph nodes)- in the above scheme.

Among the *in vitro* assays covering the above mentioned key events, internationally accepted are the direct peptide reactivity assay (DPRA) (Gerberick et al., 2004), the ARE−Nrf2 luciferase test method KeratinoSens™ (Emter et al., 2010), and the human cell-line activation test (h-CLAT) (Ashikaga et al., 2006). More tests are presently under validation (Urbisch et al., 2015; Reisinger et al., 2015). In parallel with the development of *in vitro* assays, predictive (Quantitative) Structure-Activity Relationships (QSAR) approaches have been presented as well (Teubner et al., 2013; Dearden et al., 2015). Many of these studies are aimed at modeling the results of the LLNA assay, because of its central role in the present regulatory context.

In this paper we exploit the availability of large compilations of data generated by present research, to further investigate the predictive models and to better highlight the relationships between the various *in vivo*, *in vitro* and QSAR approaches.

2. Data and methods

A large compilation of data has been made available recently as supplementary material to (Urbisch et al., 2015). This includes data from human observations and the *in vivo* LLNA assay, as well as from various *in vitro* assays. The data are presented both in their quantitative form and as summary -/+. The *in vivo* data are also given as ranks of potency.

Data are also included in the QSAR Toolbox, where they are provided as information to be used in prediction exercises (e.g., Read-Across). These data are only partially overlapping with those in the Urbisch et. compilation. The QSAR Toolbox, developed by OECD in collaboration with the European Chemical Agency (ECHA), is a standalone software application for filling gaps of (eco)toxicity data that are needed to assess the potential hazards of substances (Benigni, 2014). Version 3.3.5 is freely available on the OECD website http://www.oecd.org/chemicalsafety/risk-assessment/theoecdqsartoolbox.htm. For most of the tests, the Toolbox contains subsets of the data in Urbisch et al. However, it contains a larger collection of LLNA data, and data from GPMT and an *in vitro* Peptide Adducts test (Natsch and Gfeller, 2008), which are absent in the Urbisch et al. compilation.

Besides experimental data, the Urbisch et al., compilation contains also an ALERT for the mechanistic activity class of skin sensitizers. This ALERT was built based on results from the application of two profilers contained in the QSAR Toolbox ("Protein binding by OECD", "Protein binding by OASIS v1.2") These results were fine-tuned through expert judgement (with an eye also to the experimental results of *in vivo* systems) (Urbisch et al., 2015). This published ALERT −without any changes-was used in the QSAR analyses of the present paper.

The statistical analyses of the present paper were performed with the SAS statistical software (http://www.sas.com/en_us/software/analytics/stat.html). The methods are Regression, Canonical Discriminant, and Principal Component analyses. A very clear introduction to these methods can be found at the StatSoft, free-access website: http://www.statsoft.com/Textbook. In brief, the general purpose of multiple regression is to learn more about the relationship between several independent or predictor variables and a dependent or criterion variable. This information can be used to build a regression equation of the form, e.g.,: Toxic potency = a x Var1 + b x Var2 + …

Principal Component Analysis (PCA) points to the main trends in data consisting of many variables, separates the underlying "independent effects", estimates their relative quantitative contribution, and provides graphical displays of all the information in a few dimensions.

The general goal of discriminant analysis is to build up the function (discriminant function, D(X)) of X variables (predictors) that best discriminates two or more naturally occurring groups (e.g. negatives and positives for a biological endpoint). Canonical discriminant analysis exploits the ability of factor analysis to highlight the "true dimensionality" of a problem, and to generate summary variables from the original variables. The final discriminant function is a linear combination of the predictor variables (e.g., Toxicity class = a x DPRA + b x Keratinosens + c x h-Clat). The function is calculated for each chemical to be predicted.

3. Results and discussion

3.1. Predicting LLNA and human skin sensitization with in vitro assays

An updated, large database of human, animal, *in chemico* and *in vitro* data has been published recently by (Urbisch et al., 2015); in this context, the Authors have also analyzed the predictivity of the *in vitro* alternative methods towards human and animal (LLNA) results. Thus, it is interesting to start the present analysis by showing the main results of (Urbisch et al., 2015).

For an easier visualization, we have plotted the results tabulated in the Urbisch' paper as a Receiver Operating Characteristics (ROC) graph. The X-axis reports the False Positive Rate (i.e., 1 − Specificity), and the Y-axis reports the True Positive Rate (i.e., Sensitivity). The perfect prediction system is at the top, left corner of the graph, whereas the diagonal line represents random results (Swets, 1988).

Fig. 1 displays the prediction of the LLNA results by several *in vitro* systems, and batteries of systems on the largest data set available (n = 213). All data are qualitative (-/+). The battery '2 out of 3' model includes the DPRA, KeratinoSens™, and h-Clat tests used in combination. Fig. 1 shows that the '2 out of 3' battery approach is closest to the "perfect prediction corner". Among the individual assays, DPRA has the best overall predictivity. It should be noted that the DPRA test is a combination of two types of experiments, i.e., measurement of the depletion of cysteine and of lysine, separately. It is a probe for interactions of the chemicals with