douglasconnect
working communities

An Integrated Testing Strategy for Quantitative Skin Sensitization Risk Assessment by Integration of Multiple In Vitro Assays Using Bayesian Networks and Accounting for Bioavailability

Ahmed Abdelaziz Sayed, Maja Brajnik, Daniel Bachler, Joh Dokler, Thomas Exner and Barry Hardy
Douglas Connect GmbH, Technologiepark Basel, Hochbergerstrasse 60C, CH-4057 Basel, Switzerland
www.douglasconnect.com

Introduction

Allergic contact dermatitis (aka., skin sensitization) accounts for 20% of all contact dermatitis cases [1] and hence has an estimated annual cost up to $200 million [2]. It is also a public health problem, responsible for more than seven million outpatient visits annually. Currently, there are more than 3700 substances that are identified as contact allergens.

Traditionally, skin sensitization hazards are assessed using *in vivo* animal experiments. However, the legislation is increasingly put in place to encourage the replacement of such experiments with non-animal methods. Cosmetic products in the European Union are now banned from using animals for hazard testing. A prohibition is also in place for the sales of products that have been previously tested on animals or that contain ingredients which have undergone such testing following the ban [3]. Multiple tests have been suggested to reflect the current mechanistic understanding of the processes (see figure 1) leading to skin sensitization [5] and many framework proposals were reported to guide classification decisions with alternative non-animal data.

The current work implements an integrated testing strategy approach previously described by Jaworska et al [6]. The Integrated Testing Strategy (ITS-3) utilizes a Bayesian network (see figure 2) to assesses skin sensitization potency by combining information from three validated alternative assays, DPRA, KeratinoSens and h-CLAT as well as *in silico* predictions for bioavailability to provide a quantitative estimation for skin sensitization potency (a stimulation index = 3; EC3) across a four-category system (Non-, weak, moderate or strong sensitizers).

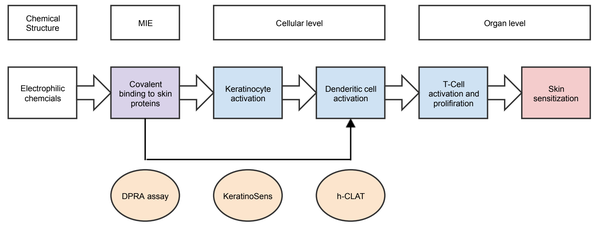

Figure 1: Diagram depicting the current understanding of the skin sensitization adverse outcome pathway as published on AOP-wiki [4].

Methodology

The prediction of chemical potency for sensitization is first calculated in the form of a probability distribution over the four potency classes. Such probability distribution is then transformed into Bayes factors to remove prediction bias from the training set distribution and to assess uncertainty giving an objective measure for judging confidence in prediction (details below). The network training is based on a training set of 147 chemicals and a test set of 60 chemicals for which the complete set of *in vivo* and *in vitro* data are available.

Value of Information (VoI)

The one-step look-ahead hypothesis was used to calculate the VoI from all network variables which were therefore ranked according to their importance in the decision process as suggested by Jaworska et al. [6] and is used as the methodology to guide testing [7] (i.e., by suggesting the variable (*in vitro* or *in silico*) for which its knowledge would maximize information gain about skin sensitization).

Correction for Michael acceptors

The posterior probability of chemicals are corrected [6] to reflect the anti-inflammatory effect of Michael acceptors [8]. The decision of whether a compound is considered a direct Michael acceptor for which correction is needed is based on SMARTS patterns. Risk assessors can override the decision.

Applicability domain

Chemical and *in vitro* assays applicability domain were taken into account. If solubility is too low for a certain assay to be indicative, such assay is considered unreliable and its values are neglected in the decision-making process. Likewise, chemicals with 100% fraction ionized are not suitable for *in vitro* assays.

Confidence estimation

The calculated Bayesian factors (BF) serve to correct the imbalance in chemical representation in the training set as well as provide a quantitative assessment of the confidence in prediction. The threshold for strength of evidence suggested by Goodman [9] were used.

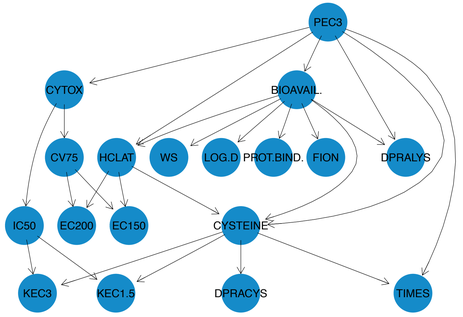

Figure 2: Diagram representing the Bayesian network as implemented in CRAN R. The network structure is based on expert knowledge [6] and reflects the current mechanistic understanding of the skin sensitization process. The network combines *in silico* models for bioavailability with three validated in vitro assays to assess the skin sensitization potency and can compensate for any missing variables.

Results

Accuracy comparison

The developed Bayesian network closely matched the performance of the published approach [4]. The overall accuracy of the network across 4 classes is 80% with class 3 showing the lowest accuracy (68%) and class 1 is the highest (95%) (see Table 1).

Effect of prediction confidence estimation using Bayesian factors

Taking prediction confidence into account highly improved the accuracy of prediction. For the most confident prediction (BF > 30), the accuracy exceeded 98%, for moderate confidence (BF between 3 and 30), the accuracy was 71% while the accuracy for the least confident predictions (BF < 3) was 60%. For a four-class classification problem, random accuracy is 25%.

Class	Jaworska et al.	DC ITS SkinSens
C1	92%	95%
C2	82%	77%
C3	70%	68%
C4	72%	79%
overall	79.6%	79.6%

Confidence class	Strong (BF >30)	Moderate (BF 3-30)	Weak (BF <3)
Correct	53	55	9
Misclassified	1	23	6
Total	54	78	15
Accuracy	**0.98**	**0.71**	**0.60**

Table 1 (left): Per-class accuracy of the developed Bayesian network (DC ITS SkinSens) in this work as compared to the accuracy of the network published in literature [4] .
Table 2 (above): The accuracy of the developed network across different confidence threshold (see confidence estimation section)

Implementation

The Bayesian network was implemented as a web application deployed on https://its.douglasconnect.com and is freely available to risk assessors. The application can also be deployed internally for interested partners and integrated with existing workflows (Garuda, KNIME, Pipleline Pilot, inter alia.).
The deployed web application allows risk assessors to use a web-interface to perform safety assessment for skin sensitization in an interactive manner. As structures are being sketched or entered as SMILES strings, the application calculates *in silico* parameters for solubility, lipophilicity, partition coefficient and protein binding. It provides instant feedback if chemicals are thought to be out of the experiments' applicability domain and suggests next experiment to be performed in order to achieve the highest value of information. The robust architecture (figure 4) allows the integration of the tool into workflow systems such as KNIME and Pipeline Pilot through he use of application programming interfaces (APIs) or in custom applications.

References and More Information

1. Coman, G.; Zinsmeister, C.; Norris, P. Occupational Contact Dermatitis. *Allergy, Asthma, Clin. Immunol.* **2008**, *4*, 59–65.
2. CDC - Skin- Occupational Dermatoses Slides 6 to 10 - NIOSH Workplace Safety and Health Topic https://www.cdc.gov/niosh/topics/skin/occderm-slides/ocderm2.html (accessed Jan 18, 2017).
3. European Union (2009) Regulation (EC) No 1223/2009 of the Euro- pean Parliament and of the Council of 30 November 2009 on cosmetic products. OJL 342(59):59–209
4. Adverse Outcome Pathway Wiki (AOP-Wiki) - Covalent Protein binding leading to Skin Sensitisation https://aopwiki.org/aops/40#taxonomic_applicability (accessed Jan 18, 2017).
5. Mehling A, Eriksson T, Eltze T, Kolle S, Ramirez T, Teubner W, van Ravenzwaay B et al (2012) Non-animal test methods for predicting skin sensitization potentials. Arch Toxicol 86(8):1273–1295
6. Jaworska, J. S.; Natsch, A.; Ryan, C.; Strickland, J.; Ashikaga, T.; Miyazawa, M. Bayesian Integrated Testing Strategy (ITS) for Skin Sensitization Potency Assessment: A Decision Support System for Quantitative Weight of Evidence and Adaptive Testing Strategy. *Arch. Toxicol.* **2015**, *89*, 2355–2383.
7. Kjaerulff UB, Madsen AL (2013) Bayesian networks and influence diagrams: a guide to construction and analysis, 2nd edn. Springer, New York
8. 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 S NAr- and S N2-reactive chemicals. Chem Res Toxicol 24(11): 2018–2027
9. Goodman SN (1999) Toward evidence-based medical statistics. 2: the Bayes factor. Ann Intern Med 130(12):1005–1013
10 ECHA registration dossier on Vanillin https://echa.europa.eu/registration-dossier/-/registered-dossier/2209/7/5/2

Case Studies

Vanillin

Vanillin (CASRN: 121-33-5) is the primary component of the extract of the vanilla bean and is used as a flavouring agent in foods, beverages, and pharmaceuticals. Animal studies in guinea pig (maximization test) [10] and mice (LLNA assay) [6] has shown the compound to be a non-sensitizer. However, TIMES *in silico* model for skin sensitization predicts the compound as a strong sensitizer (category 3 on TIMES scale). Therefore, relying only on the *in silico* prediction might unnecessarily disqualify a safe chemical. However, when entering the compound structure along with the TIMES prediction, the DC BN ITS-3 tool calculates physicochemical and bioavailability descriptors and considers all collected input to reach the conclusion that the confidence in prediction is weak (Bayes factor = 2.97; i.e., less than 3) [figure 3 A]. The algorithm then suggests performing the h-CLAT in vitro assay for maximum information gain. When the assay is performed and results fed back, the confidence of the compound being a moderate sensitizer becomes even lower (Bayes factor = 2.08) [figure 3 B]. The tool suggests performing the KeratinoSens™ assay as the next step, which leads to the final prediction becoming "non-sensitizer" but with rather weak confidence (Bayes factor = 2.72) [figure 3 C] and the algorithm requesting the DPRA assay results to improve the confidence. Only after performing the third *in vitro* assay (i.e., DPRA) does the confidence become substantial (Bayes factor = 3.76) [figure 3 D] that the compound is indeed a non-sensitizer despite the *in silico* prediction judging otherwise. The Bayesian network was able to take all evidence into account to reach an accurate final conclusion about the safety of the compound as non-sensitizer despite misleading initial *in silico* prediction. This was also the case with other compounds such as Saccharin and Benzen 1-methoxy-4-methyl-2-nitro.

3-Decen-5-one 4-methyl- (3E)

The compound is predicted by the Bayesian network as a weak sensitizer with substantial confidence (Bayes factor = 4.84; i.e., 3< confidence< 30). This indeed matches the LLNA assay results. However, risk assessors might want to understand the confidence of this particular prediction by diving deeper into the probability graph provided in the detailed report for compound safety. The 50th percentile corresponds to the Bayesian network best estimate of the predicted outcome. While the 50th and 60th percentiles fall into the weak-sensitizer category (rather on the higher threshold), the 80th and 90th percentiles fall under the moderate-sensitizing category. Risk assessors can transparently judge (according to the case at hand, exposure, or other factors) how conservative they want their decision. The detailed report includes all the charts accompanied by tables that shows the corresponding EC3 and PEC3 values.

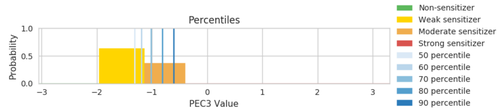

Figure 4: Diagram depicting the different percentile probabilities for pEC3 values predicted for 3-Decen-5-one 4-methyl- (3E). The horizontal lines represents (50th, 60th, 70th, 80th and 90th) probability percentiles.

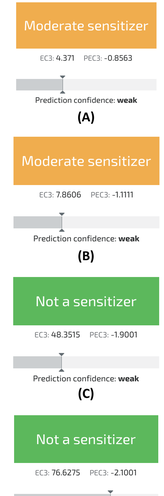

Figure 4 A-D: Diagram showing the evidence building for vanillin with conducted experiments. The algorithm guides the risk assessor on which experiment to conduct next in order to reach a high confidence in prediction (D)

Conclusions and Future Directions

Advantages of the Bayesian network approach: (1) It tolerates missing information and conveys a probabilistic hypothesis of skin sensitization based on accumulative evidence from data. (2) It assesses the uncertainty in prediction given the input data using Bayes Factors. (3) It directs consequent testing by value of information calculations, i.e. suggesting the experiments to be conducted to achieve maximum information gain reducing uncertainty in prediction. (4) Risk assessors are presented with different levels of confidence for decision making.

Next steps in collaboration with industrial partners and research consortia:
(1) Together with partners from the cosmetics, and fragrances industries, we are expanding the applicability domain of the models using internal data.
(2) We are extending the application to include more defined approaches with different machine-learning algorithms and more *in vitro* assays and *in silico* models. An approach is being developed for guiding risk assessors to suitable defined approaches for their chemicals.
3) Investigating further endpoints using the Bayesian network approach and integrating biological knowledge from adverse outcome pathway networks.

Acknowledgements

We would like to thank Dr. Joanna Jaworska for her support of the project and the Cheminformatics community for providing open tools that made such developments possible. This work was funded by the Eurostars-2™ Programme under Project ID: 9698

ToxHQ