Addressing the challenge of making negative predictions for skin sensitisation

Martyn L. Chilton, Donna S. Macmillan, Richard V. Williams

Lhasa Limited, Granary Wharf House, 2 Canal Wharf, Leeds, LS11 5PS, UK

Abstract number – 3138

1. Introduction

Why are negative predictions important?

In silico models are expected to be relied upon increasingly in the future, with a number of recent regulations from around the globe promoting alternative methods, and in some cases even banning the use of traditional *in vivo* testing. However, a user of an *in silico* model needs to be particularly confident in a negative prediction, as an incorrect prediction could potentially contribute to human safety being compromised.

Can negative predictions be made for skin sensitisation?

The question of how to predict the absence of bacterial *in vitro* mutagenicity was recently addressed using an *in silico* fragmentation approach[1]. This research project aimed to extend this methodology to make negative predictions of human skin sensitisation, being aware of the opportunities and challenges that this particular toxicity endpoint afforded (Table 1).

Opportunities	Challenges
• Adverse outcome pathway with a single reactivity-driven molecular initiating event • Well-developed models for making positive *in silico* predictions (e.g. Derek Nexus contains 90 alerts for skin sensitisation)[2]	• Prediction of an adverse outcome in humans using *in vivo* data from multiple assays and species • The quantity of *in vivo* data available in the public domain

Table 1. Challenges and opportunities expected when making negative predictions of human skin sensitisation.

2. How negative predictions are made

Constructing the skin sensitisation negative dataset

High quality human data consisting of expert-derived potency categorisations[3-6] was combined with animal data extracted from the Vitic Nexus database[7]. The chemical and biological data were curated before the various assays were categorised according to their relevance and reliability in predicting skin sensitisation in humans (Figure 1). An overall call was assigned to each chemical using a hierarchical approach, with human data considered first where available, followed by standard animal, non-standard animal and finally positive other animal data. The final dataset contained 1346 sensitisers and 1420 non-sensitisers.

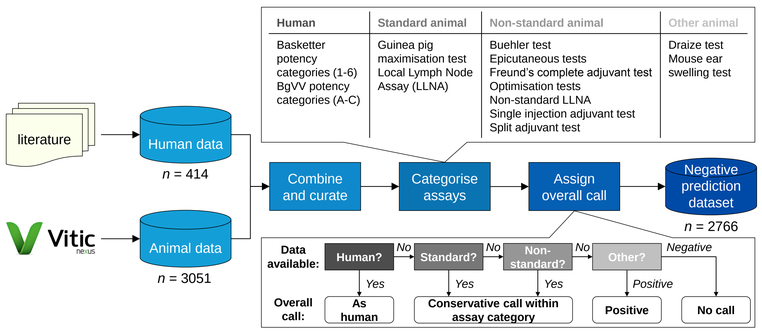


Figure 1. Construction of the skin sensitisation negative prediction dataset.

Making negative predictions within an expert knowledge-based system

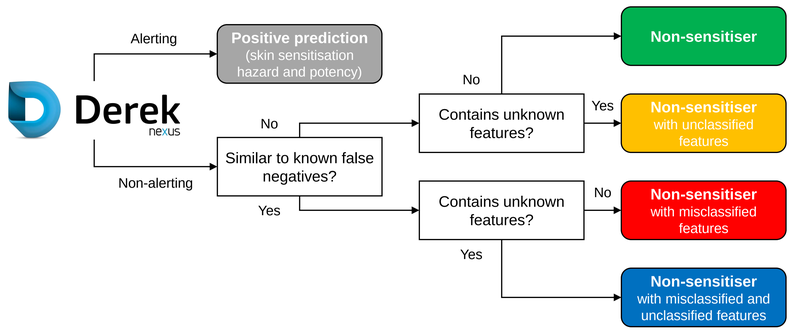


Figure 2. How negative predictions can be made within an expert knowledge-based system.

The negative prediction approach seeks to answer two questions of a non-alerting chemical in Derek Nexus: (i) is it similar to known false negatives, and (ii) does it contain any unknown features? (Figure 2). These are addressed by comparing the query chemical to the skin sensitisation negative prediction dataset. Features which are only found in known false negatives within the dataset are highlighted as **misclassified features**, whereas those which are not present in the dataset are flagged as **unclassified features**.

3. Performance in 5-fold cross validation

The skin sensitisation negative prediction dataset was randomly split into 5 folds to carry out a cross-validation. The mean performance figures across the 5 test sets are shown below.

How often does each negative prediction outcome occur?

The majority of non-alerting chemicals (**79%**, Figure 3) were classified as non-sensitisers. Misclassified or unclassified features were only identified **8%** and **12%** of the time, respectively. **Fewer than 1%** of chemicals contained both sets of features.

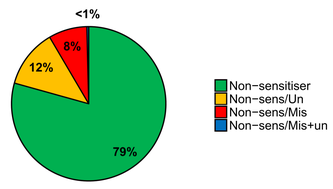
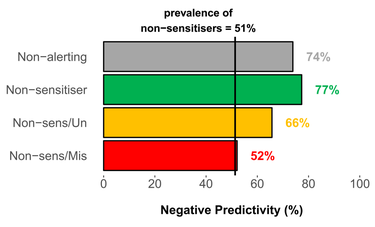


Figure 3. Distribution of negative prediction outcomes.

How often is each negative prediction outcome correct?

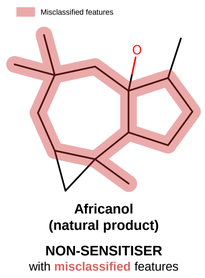


Non-alerting chemicals which contained no misclassified or unclassified features were correctly predicted to be non-sensitisers more often than the test set as a whole (**77%** cf. 74%, Figure 4). Conversely, misclassified and unclassified chemicals were both correctly predicted less often than the overall test set (**66%** and **52%** cf. 74%).

Figure 4. Negative predictivity for all non-alerting chemicals and for each prediction outcome.

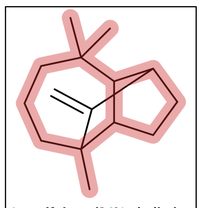
4. Expert review of resulting predictions

Case study 1: confidence in the negative prediction increases



Africanol (natural product)

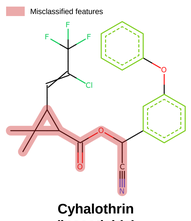
NON-SENSITISER with misclassified features



Longifolene (34% similar)
Human positive (weak)[3]
LLNA positive (moderate)[8]
Possible prehapten[9]

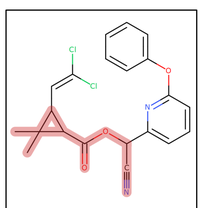
- Africanol contains a misclassified feature, derived from the known false negative longifolene.
- However, longifolene may act as a prehapten[9], probably by producing a sensitising allylic hydroperoxide[10].
- As africanol lacks an allyl group, confidence in the negative prediction is likely to *increase*.

Case study 2: confidence in the negative prediction decreases



Cyhalothrin (insecticide)

NON-SENSITISER with misclassified features



Fenpyrithrin (63% similar)
Guinea pig split adjuvant test positive (5/5 animals sensitised)[11]

- Cyhalothrin also contains a misclassified feature, derived from the known false negative fenpyrithrin.
- The two chemicals are highly structurally similar, and share the same overall scaffold.
- Given the high similarity to the false negative, confidence in the negative prediction is likely to *decrease*.

5. Conclusions

Confident negative predictions of human skin sensitisation can be made

An *in silico* fragmentation approach has been applied to make negative predictions of human skin sensitisation. The majority of **non-alerting chemicals** in the cross-validation test sets were simply predicted to be **non-sensitisers**. As this outcome was correctly predicted most often, these predictions can be treated with a higher level of confidence.

Features which increase uncertainty are highlighted for expert review

The presence of **misclassified** or **unclassified** features will decrease the confidence a user has in the negative prediction, as these outcomes were correctly predicted less often. Expert review may be able to resolve this residual uncertainty, as demonstrated in the case studies.

6. References

1. R. V. Williams et al., *Regul. Toxicol. Pharmacol.* **2016**, *76*, 79-86
2. Derek Nexus v6.0.0 (Lhasa Limited), , www.lhasalimited.org/products/derek-nexus.htm
3. A. M. Api et al., *Dermatitis* **2017**, *28*, 299-307
4. D. A. Basketter et al., *Dermatitis* **2014**, *25*, 11-21
5. D. Kayser, E. Schlede (Eds.), *Chemikalien und Kontaktallergie – Eine bewertende Zusammenstellung*, **2001**, Verlag Urban & Vogel, Munchen
6. E. Schlede et al., *Toxicology* **2003**, *193*, 219-259
7. Vitic Nexus v2.6.1 (Lhasa Limited), www.lhasalimited.org/products/vitic-nexus.htm
8. A. Natsch et al., *Toxicology In Vitro* **2007**, *21*, 1220-1226
9. SCCS (Scientific Committee on Consumer Safety), *Opinion on fragrance allergens in cosmetic products*, **2012** http://ec.europa.eu/health/scientific_committees/consumer_safety/docs/sccs_o_102.pdf
10. A.-T. Karlberg et al., *Contact Dermatitis* **2013**, *69*, 323-334
11. K. S. Rao et al., *Drug and Chemical Toxicology* **1981**, *4*, 331-351