Regulatory Toxicology and Pharmacology 76 (2016) 79–86

Contents lists available at ScienceDirect

Regulatory Toxicology and Pharmacology

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

# It's difficult, but important, to make negative predictions

Richard V. Williams [a, *], Alexander Amberg [b], Alessandro Brigo [c], Laurence Coquin [a], Amanda Giddings [d], Susanne Glowienke [e], Nigel Greene [f, 1], Robert Jolly [g], Ray Kemper [h], Catherine O'Leary-Steele [a], Alexis Parenty [e], Hans-Peter Spirkl [b], Susanne A. Stalford [a], Sandy K. Weiner [i], Joerg Wichard [j]

[a] Lhasa Limited, Granary Wharf House, 2 Canal Wharf, Leeds, LS11 5PS, UK
[b] Sanofi-Aventis Deutschland GmbH, R&D DSAR/Preclinical Safety FF, Industriepark Hoechst, Bldg. H823, D-65926 Frankfurt am Main, Germany
[c] Roche Pharmaceutical Research & Early Development, Pharmaceutical Sciences, Roche Innovation Center Basel, Grenzacherstrasse 124, CH-4070 Basel, Switzerland
[d] GlaxoSmithKline Pre-Clinical Development, Park Road, Ware, Hertfordshire SG12 0DP, UK
[e] Novartis Pharma AG, Pre-clinical Safety, Werk Klybeck, CH-4057 Basel, Switzerland
[f] Compound Safety Prediction, Pfizer Global Research & Development, 558 Eastern Point Road, Groton, CT 06340, USA
[g] Health/Safety/Environmental, Lilly Research Laboratories, Indianapolis, IN, USA
[h] Vertex Pharmaceuticals Incorporated, 50 Northern Ave, Boston MA, USA
[i] Janssen Research & Development, 1400 McKean Road, Spring House, PA 19477, USA
[j] Bayer Pharma AG, Investigational Toxicology, Muellerstr. 178, S 116, D-13353, Berlin, Germany

ARTICLE INFO

*Article history:*
Received 27 November 2015
Received in revised form
13 January 2016
Accepted 14 January 2016
Available online 16 January 2016

*Keywords:*
Negative predictions
(Q)SAR
Expert system
*In silico* toxicology
Expert assessment
ICH M7

ABSTRACT

At the confluence of predictive and regulatory toxicologies, negative predictions may be the thin green line that prevents populations from being exposed to harm. Here, two novel approaches to making confident and robust negative *in silico* predictions for mutagenicity (as defined by the Ames test) have been evaluated. Analyses of 12 data sets containing >13,000 compounds, showed that negative predictivity is high (~90%) for the best approach and features that either reduce the accuracy or certainty of negative predictions are identified as misclassified or unclassified respectively. However, negative predictivity remains high (and in excess of the prevalence of non-mutagens) even in the presence of these features, indicating that they are not flags for mutagenicity.



## 1. Introduction

Negative predictions pose difficult questions for both developers and users of *in silico* toxicology prediction tools.

The question for model developers is 'how can we predict the absence of a signal,' given that activity (be it genotoxicity, sensitisation, hepatotoxicity or even endpoints outside of toxicology) can be considered a property of compounds that differ from the background. That is to say, we have a prior expectation that compounds will be inactive unless they contain a feature (or exhibit a property) that causes or induces toxicity. This is exemplified in the language we use when we ask (and try to identify) 'what is it about a compound that **causes** activity?' This is not intended to be a definition, merely a summary of the prevailing thinking.

The question for model users is clearer, being: 'can I trust this negative prediction for my compound'? In itself, this question is an oversimplification as the situation is more akin to 'is the absence of a positive prediction sufficient evidence for a negative prediction for my compound?'

In addition to being difficult, negative predictions are also important. We are advancing into times where alternative methods for toxicity prediction are being used more than ever, as illustrated through the rise in PubMed citations for predictive toxicology (Fig. 1).

This is driven by the principle that we should move away from

* Corresponding author.
*E-mail address:* richard.williams@lhasalimited.org (R.V. Williams).
[1] Current affiliation: Predictive Compound Safety and ADME, AstraZeneca, 35 Gatehouse Drive, Waltham, Massachusetts 02451, USA.

http://dx.doi.org/10.1016/j.yrtph.2016.01.008
0273-2300/

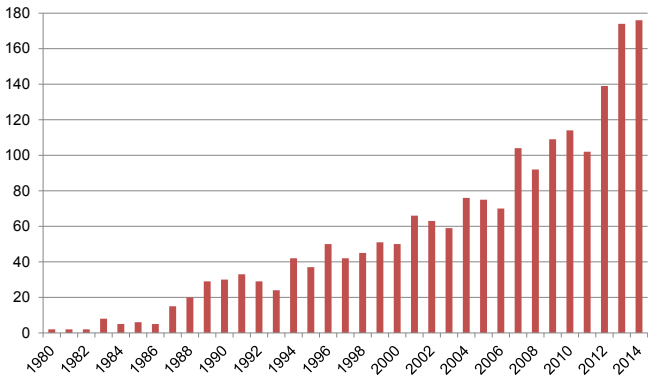


**Fig. 1.** Pubmed citations for predictive toxicology (accessed 23/02/2015).

the use of animal testing for human risk assessments (Annys et al. 2014), the potential that alternatives may be better models for the true human situation than animal models (Chapman et al. 2013) and the recognition that in some situations conducting *in vitro* or *in vivo* tests is not practical. In the latter case, *in silico* models are considered to be 'good enough' for specific applications, e.g. when used alone in the assessment of potential mutagenic impurities in pharmaceuticals (Dobo et al. 2006; Mueller et al., 2006) or used in combination with read-across and TTC (Threshold of Toxicological Concern) for pesticide metabolites (European Food Standards Agency, 2012).

These drivers are captured in some of the strategies (e.g. Tox21 (National Research Council, 2007)), guidelines (e.g. ICH M7 (International Conference on Harmonization, 2014)) and regulations (e.g. REACH (European Chemicals Agency, 2012)) published over the last ten years. The logical conclusion to this is that *in silico* predictions will increasingly be used as the primary line of defence that prevents populations from being exposed to a hazard.

In cases where model users are presented with a positive prediction, two viable options are available:

- Accept positive prediction and reduce or control exposure to the predicted hazardous compound
- Explore the true risk posed by the predicted hazardous compound (e.g. by running an *in vitro* or *in vivo* assay or examining in more detail exposure levels and thresholds of concern)

Whilst this may increase the burden on, for example, a company attempting to bring a product to a market, it is likely to ensure the safety of exposed populations. By contrast, a negative prediction presents a bigger challenge, because of its significance. If this is to be taken in lieu of an experimental result then we assume the negative prediction translates to a lack of hazard and therefore lack of risk. Thus, an erroneous negative prediction could result in exposure of a population to a hazardous substance with potential deleterious effects on public health. Therefore these predictions are subjected to a tighter scrutiny (Powley, 2015).

Historically *in silico* systems predict inactivity by inferring from the true evidence, which is a lack of a positive signal for activity. Often, for models derived through machine-learning, this is combined with an estimation of applicability domain. Thus, compounds are predicted inactive if no reasons for activity are found, and the chemical is considered within the applicability domain of the model.

Applicability domains themselves may complicate the interpretation of *in silico* toxicity predictions and have been extensively reviewed. Suffice to say here that (i) many approaches for determining applicability domains are possible (Jaworska et al., 2005) potentially producing different results (Sahigara et al., 2012), (ii) the outcome of applicability domain assessments are not always meaningful (Ellison et al., 2011) and (iii) pushing query compounds outside of a theoretical applicability domain can increase the number of compounds that need to be tested (*in vitro* or *in vivo*), significantly reducing the value of an *in silico* screen (Jolly et al., 2015) unless the results are used to build future models.

In this work, we explore alternative or additive methods that can be used to increase the robustness of, and confidence in, negative predictions for activity in the bacterial reverse mutation assay (commonly referred to as the Ames test). Firstly, by defining regions around alerting space (so-called 'predictive space') then by evaluating similarity to known (Ames test) mutagens whose activity is incorrectly predicted by *in silico* systems (herein referred to as false negatives).

## 2. Material and methods

### 2.1. Predictive space work

The Derek Nexus (version 2.0, the latest available at the time) knowledge base was used as a starting point for this work. Derek Nexus is an expert knowledge base system for toxicity prediction, containing rules derived from both public and proprietary data (e.g. as described in Elder et al., 2015). Each bacterial, *in vitro* mutagenicity alert in the knowledge base was examined by a scientist with expertise in mutagenicity alert development. The patterns encoding the SAR for each alert were modified using the Derek Knowledge Editor (version 2.0) if they contained features that were implemented to prevent the pattern being activated by non-mutagenic compounds (so-called exclusion patterns). Such features were removed and the resultant 'predictive space' stored within a modified knowledge base. Thus, each bacterial, *in vitro* mutagenicity alert in Derek had a corresponding region of predictive space.

The ability of predictive space to make accurate negative predictions was assessed using three Ames test data sets (Table 1).

(*Benchmark data set*) – originally published by Hansen et al. (2009), later curated internally as reported in Sherhod et al. (2014).
(*Lhasa Vitic intermediates data set*) – provided by the Lhasa Limited intermediates data sharing group, extracted from Vitic Nexus.
(*ECHA chem data set*) – extracted from European Chemical Agency registration dossiers (available via www.echemportal.org).

### 2.2. Evaluating similarity to known false negatives

A negative prediction reference set was compiled from 5 publically available Ames test data sets and 1 data set donated by FDA CFSAN with Ames test data (Table 2). Most of these data sets were initially used in Derek for alert validation, and later all were used for the development of new and modified alerts. It should be noted,

**Table 1**
Data sets used for evaluation of predictive space.

| | Mutagens | Non-mutagens |
|---|---|---|
| Benchmark data set | 3503 | 3009 |
| ECHA chem data set | 240 | 2271 |
| Vitic intermediates data set* | 279 | 600 |