Regulatory Toxicology and Pharmacology 79 (2016) 12–24

Contents lists available at ScienceDirect

Regulatory Toxicology and Pharmacology

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

Systematically evaluating read-across prediction and performance using a local validity approach characterized by chemical structure and bioactivity information

Imran Shah [a, *], Jie Liu [b, c], Richard S. Judson [a], Russell S. Thomas [a], Grace Patlewicz [a]

[a] National Center for Computational Toxicology, Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, NC 27711, USA
[b] Department of Information Science, University of Arkansas at Little Rock, AR 72204, USA
[c] Oak Ridge Institute for Science Education Fellow, National Center for Computational Toxicology, Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, NC 27711, USA

ARTICLE INFO

Article history:
Received 25 September 2015
Received in revised form
20 April 2016
Accepted 3 May 2016
Available online 9 May 2016

Keywords:
Read-across
Nearest neighbors
Local validity domains
(Q)SAR
KNN
Bioactivity
ToxCast

ABSTRACT

Read-across is a popular data gap filling technique within category and analogue approaches for regulatory purposes. Acceptance of read-across remains an ongoing challenge with several efforts underway for identifying and addressing uncertainties. Here we demonstrate an algorithmic, automated approach to evaluate the utility of using *in vitro* bioactivity data ("bioactivity descriptors", from EPA's ToxCast program) in conjunction with chemical descriptor information to derive local validity domains (specific sets of nearest neighbors) to facilitate read-across for up to ten *in vivo* repeated dose toxicity study types. Over 3239 different chemical structure descriptors were generated for a set of 1778 chemicals and supplemented with the outcomes from 821 *in vitro* assays. The read-across prediction of toxicity for 600 chemicals with *in vivo* data was based on the similarity weighted endpoint outcomes of its nearest neighbors. The approach enabled a performance baseline for read-across predictions of specific study outcomes to be established. Bioactivity descriptors were often found to be more predictive of *in vivo* toxicity outcomes than chemical descriptors or a combination of both. This generalized read-across (GenRA) forms a first step in systemizing read-across predictions and serves as a useful component of a screening level hazard assessment for new untested chemicals.

1. Introduction

1.1. Problem formulation

Read-across is a widely used technique for filling data gaps within category and analogue approaches for regulatory purposes. In the European Union, under the Registration, Evaluation, Authorization and Restriction of Chemicals (REACH) regulation (EC, 2006), a 'read-across' or category approach has been used in up to 75% of analyzed dossiers for at least one endpoint (ECHA, 2014). A similar proportion of High Production Volume (HPV) chemicals submitted to the US EPA under the Toxic Substances Control Act (TSCA) have been initially evaluated using read-across approaches (Bishop et al., 2012). Despite the broad application of read-across approaches, regulatory acceptance remains a major hurdle primarily due to the lack of objectivity and clarity about how to practically address uncertainties. Currently, read-across is largely a subjective expert judgement-driven assessment in terms of both analogue selection and data interpretation (ECETOC, 2012; Cronin et al., 2013; Patlewicz et al., 2013, 2014a).

There is a continuum of non-testing methods[1] for relating chemical structural information (and *in vitro* bioactivity) to toxicity outcomes (Tollefsen et al., 2014). A specific approach is chosen dependent on the amount of available data, the endpoint of interest, and the regulatory context. When a large set of chemicals with known toxicity are available, then supervised machine learning

[1] Non-testing approaches was a term coined during the development of the REACH Technical guidance to encompass the search and retrieval of existing information, the identification of structural alerts, to the development of (Q)SARs and the grouping of chemicals for read-across.

* Corresponding author.
E-mail address: shah.imran@epa.gov (I. Shah).

http://dx.doi.org/10.1016/j.yrtph.2016.05.008
0273-2300/

methods can be used to automatically build and objectively evaluate the quantitative structure activity relationships (QSARs) derived. On the other hand, when there are only a handful of chemicals with toxicity data, then quantitative relationships between structural descriptors and activities are more likely to be subject to "over-fitting," often rendering those QSARs unreliable. In such data-poor situations, read-across serves as a pragmatic option since experts can use prior knowledge to identify appropriate (source) analogues and use them to infer activity of a new (target) chemical. Experts can also substantiate the read-across prediction using mechanistic associations between structure and activity, an explanatory quality that is often absent in QSARs, particularly global QSARs (Tollefsen et al., 2014).

To gain regulatory acceptance, it is important to evaluate the uncertainty in read-across predictions, and to improve scientific confidence by taking advantage of new *in vitro* bioactivity data streams which have the potential to provide mechanistic information (Patlewicz et al., 2015a). We believe this may be feasible by formulating read-across predictions in the context of a local validity domain (Patlewicz et al., 2014a) that is amenable to objective evaluation using a QSAR-like framework (OECD, 2004, 2007; ECHA, 2008).

In this study, we present a novel framework to formalize read-across based on a weight-of-evidence approach that uses the activities of structurally related chemicals to infer the activity of a new (target) chemical. We describe a unified approach to systematically evaluate confidence in read-across predictions of toxicity using chemical structure, and a potential solution for reducing uncertainty using *in vitro* bioactivity data. An approach was proposed by Low et al. (2013) to predict the toxicity of a chemical by using a similarity-weighted activity of nearest neighbors (see Methods for a more detailed description). Here we propose a generalization of the Low et al. (2013) approach, called generalized read-across (GenRA), and use it to predict toxicity across structurally similar neighborhoods in the ToxCast library containing 1778 chemicals. Although GenRA is a naïve approximation of expert judgement-driven read-across, we believe it is useful for establishing a performance baseline upon which future improvements can be made. To reduce uncertainty of predictions of toxicity based on chemical structure alone, we evaluated the read-across performance by incorporating *in vitro* bioactivity descriptors based on 821 ToxCast assays. An important contribution of our work is the use of objective metrics to estimate confidence in read-across predictions of toxicity for untested target compounds using a different number of source chemicals with known bioactivity.

1.2. Challenges with acceptance of current read-across approaches

Regulatory acceptance of read-across remains a particular challenge under REACH, especially for repeated dose toxicity endpoints as evaluated within guideline 90 day, reproductive or developmental toxicity studies (Ball et al., 2014; Patlewicz et al., 2014a). It has been postulated that formulating a framework to make explicit the critical information required for a read-across, and identifying the sources of uncertainties would be a helpful step forward, thus identifying where and what uncertainties may arise and practical ways to manage them (Patlewicz et al., 2014a). Various scientific experts have attempted to do this. Examples include the ECHA read-across assessment framework (RAAF), which is intended to assist evaluators in assessing read-across justifications within REACH dossiers (de Raat, 2014; ECHA, 2015). Researchers at P&G have proposed a framework to evaluate the uncertainties associated with a read-across framework particularly for Developmental and/or Reproductive Toxicity (DART) endpoints (Blackburn and Stuard, 2014; Blackburn et al., 2015). The EU Safety Evaluation Ultimately Replacing Animal Testing (SEURAT-1) Research Initiative developed templates to document uncertainty in read-across (Schultz et al., 2015). Finally, a read-across team established under the Long Range Initiative Programme of the European Chemistry Industry Council (Cefic LRI) proposed scientific confidence considerations in the development and evaluation of read-across justifications (Patlewicz et al., 2015a).

Whilst all these efforts are helpful starting points, none have so far tackled practical ways of reducing uncertainties in a read-across prediction. The framework by Blackburn and Stuard (2014) put forward assessment factors to address uncertainties, these are appropriate for endpoints where a dose descriptor or point of departure (POD) such as a No Adverse Effect Level (NOAEL) can be established. An assessment factor could then be used to adjust the POD value being taken up in the subsequent risk assessment. However, use of the Blackburn and Stuard (2014) framework could result in a reference dose being derived (e.g., the derived no effect level (DNEL)), that is too low to be practically useful in the context of risk management. An assessment factor approach also can not address the uncertainty in a read-across for binary endpoints such as skin sensitization or mutagenicity. Patlewicz et al. (2015a) proposed two strategies in an effort to build scientific confidence in read-across for specific decision contexts, either making use of an adverse outcome pathway (AOP) (Ankley et al., 2010; Patlewicz et al., 2015b; Tollefsen et al., 2014) or through the use of bioactivity data such as that generated within ToxCast (Judson, 2010), assuming an AOP was not available or where the mechanism was insufficiently understood to develop one. Ultimately, the goal is to make use of appropriate mechanistic information to reduce the uncertainty in a read-across prediction (Patlewicz et al., 2015a).

Whilst the approaches proposed by Patlewicz et al. (2015a) may provide enhancements to current read-across, which is still very much rooted in chemical structural similarity, neither addresses the issue of read-across performance. Read-across remains an expert, judgement-driven assessment that is context dependent on the chemical and endpoint under consideration. There is presently an absence of illustrative case studies to clarify how a regulatory agency will evaluate read-across justifications, such that an objective framework to systematically evaluate read-across performance is very much needed.

Indeed one can contrast the expert driven approach typically used in read-across with the automated (Q)SAR approach, which also uses information on structurally-related chemicals, but does so in a way that is objective, and amenable to making estimates of uncertainty in predictions (Cronin et al., 2013; Patlewicz et al., 2016). Both approaches are underpinned by the notion of toxicity being a function of chemical structure, but they simply represent different regions along the continuum of non-testing approaches. Thus, the principles described for (Q)SAR application such as those outlined for REACH (ECHA, 2008) could be helpful. In particular, the framework for assessing a QSAR prediction – whether the model addresses a defined endpoint, to what extent the QSAR is scientifically valid with respect to the OECD Validation Principles (OECD, 2004, 2007) and whether the substance of interest falls within the applicability domain (ECHA, 2008; Netzeva et al., 2005) could also be relevant for read-across. For a QSAR, the applicability domain is usually extracted from the training set used to develop the model itself, and there are a number of different approaches that can be employed to derive a domain (Netzeva et al., 2005; Dimitrov et al., 2005). For SARs, the concept of an applicability domain is still evolving (see Ellison et al. (2011) for a summary of several different approaches). More recently, Patlewicz et al. (2014b) proposed the concept of a structural alert reliability within the context of the hybrid expert system Tissue Metabolism Simulator for Skin Sensitization (TIMES-SS) (Patlewicz et al., 2007,