Science of the Total Environment 636 (2018) 901–909

Contents lists available at ScienceDirect

Science of the Total Environment

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

Rapid experimental measurements of physicochemical properties to inform models and testing☆

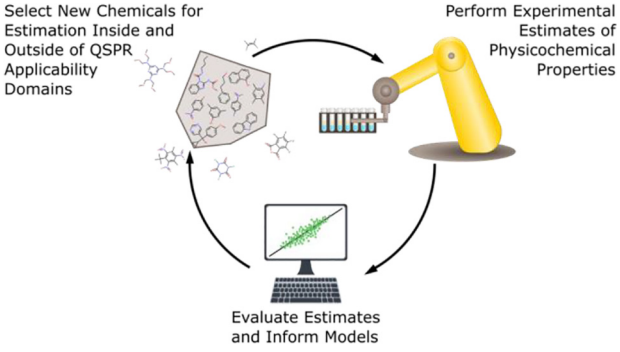
Chantel I. Nicolas [a,b,e], Kamel Mansouri [a,b,e], Katherine A. Phillips [c], Christopher M. Grulke [b], Ann M. Richard [b], Antony J. Williams [b], James Rabinowitz [b], Kristin K. Isaacs [c], Alice Yau [d], John F. Wambaugh [b,*]

[a] ScitoVation, LLC 6 Davis Drive, Durham, NC 27703, USA
[b] National Center for Computational Toxicology, Office of Research and Development, US EPA, Research Triangle Park, NC 27711, USA
[c] National Exposure Research Laboratory, Office of Research and Development, US EPA, Research Triangle Park, NC 27711, USA
[d] Southwest Research Institute, San Antonio, TX 78238, USA
[e] Oak Ridge Institute for Science and Education, Oak Ridge, TN 37831, USA

HIGHLIGHTS

- High-throughput measurements of five physicochemical properties for 200 compounds were attempted.
- New data are now available for optimizing physicochemical property QSPR models.
- Data from rapid property measurements will help reduce uncertainty in QSAR models that inform chemical risk assessment.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 20 February 2018
Received in revised form 19 April 2018
Accepted 20 April 2018
Available online xxxx

Editor: D. Barcelo

Keywords:
Physicochemical properties
Environmental chemicals
Predictive modeling
QSAR
Chemical features

ABSTRACT

The structures and physicochemical properties of chemicals are important for determining their potential toxicological effects, toxicokinetics, and route(s) of exposure. These data are needed to prioritize the risk for thousands of environmental chemicals, but experimental values are often lacking. In an attempt to efficiently fill data gaps in physicochemical property information, we generated new data for 200 structurally diverse compounds, which were rigorously selected from the USEPA ToxCast chemical library, and whose structures are available within the Distributed Structure-Searchable Toxicity Database (DSSTox). This pilot study evaluated rapid experimental methods to determine five physicochemical properties, including the log of the octanol:water partition coefficient (known as log(K_{ow}) or logP), vapor pressure, water solubility, Henry's law constant, and the acid dissociation constant (pKa). For most compounds, experiments were successful for at least one property; log(K_{ow}) yielded the largest return (176 values). It was determined that 77 ToxPrint structural features were enriched in chemicals with at least one measurement failure, indicating which features may have played a role in rapid method failures. To gauge consistency with traditional measurement methods, the new measurements were compared with previous measurements (where available). Since quantitative structure-activity/property relationship (QSAR/QSPR) models are used to fill gaps in physicochemical property information, 5 suites of QSPRs were evaluated for their predictive ability and chemical coverage or applicability domain of new experimental

☆ Disclaimer: The views expressed in this publication are those of the authors and do not necessarily represent the views or policies of the U.S. Environmental Protection Agency. Reference to commercial products or services does not constitute endorsement.

* Corresponding author at: 109 T.W Alexander Dr., NC 27711, USA.
E-mail address: Wambaugh.john@epa.gov (J.F. Wambaugh).

https://doi.org/10.1016/j.scitotenv.2018.04.266
0048-9697/Published by Elsevier B.V.

measurements. The ability to have accurate measurements of these properties will facilitate better exposure predictions in two ways: 1) direct input of these experimental measurements into exposure models; and 2) construction of QSPRs with a wider applicability domain, as their predicted physicochemical values can be used to parameterize exposure models in the absence of experimental data.

Published by Elsevier B.V.

1. Introduction

Physicochemical properties such as the log of the octanol:water partition coefficient (log(K_{ow}) or logP), and vapor pressure (VP) play a critical role in addressing many aspects of a chemical's behavior including in drug discovery (Lipinski et al., 1997; Martin et al., 2010; Arnott et al., 2012; Isaacs et al., 2016; Phillips et al., 2017), migration through the environment and body (Arnot et al., 2012; Arnot, 2006; MacLeod, 2010; Nichols et al., 2013; Peyret et al., 2010; Rosenbaum et al., 2008; Schmitt, 2008; MacLeod, 2014), and potential impact on human health and the environment (Cronin et al., 2003; Sanderson et al., 2003; Price et al., 2009; Mackay et al., 2009). In order to estimate the environmental risk posed by such chemicals, cheminformatics tools and predictive models rely on physicochemical properties to predict important aspects such as toxicity (Cronin et al., 2003; Sanderson et al., 2003; Price et al., 2009; Mackay et al., 2009), toxicokinetics (Peyret et al., 2010; Schmitt, 2008; Wambaugh et al., 2015), and exposure(Arnot et al., 2012; Arnot, 2006; Rosenbaum et al., 2008). However, there are tens of thousands of manufactured chemicals that may find their way into living organisms and the environment (USGAO, n.d.; USGAO, 2013; Schymanski et al., 2015; Rager et al., 2016; Park et al., 2012) but also have few or no physicochemical property data. This lack of data is particularly problematic as physicochemical properties govern how chemicals: 1) affect the biosphere (i.e., physiological and pathological effects); and 2) emit from or pass through the lithosphere (soil), hydrosphere (water), and atmosphere (air) to arrive at biological sites of exposure. Advancements in computational toxicology methods support on-going efforts to develop rapid toxicity information to inform the anthroposphere (human evaluation and regulation) in a decision-making context (Arnot et al., 2008; Nabholz, 1991; Walker et al., 2002; Swanson et al., 1997; Reuschenbach et al., 2008; European Chemicals Agency, 2012; Benfenati et al., 2011; Netzeva et al., 2005). The USEPA's Exposure Forecasting (ExpoCast) project (Cohen Hubal et al., 2010) relies upon robust physicochemical property information in order to develop high-throughput exposure and toxicokinetic models. In lieu of measured data, property prediction models can potentially be used, but the validity and relevance of these models need to be assessed as new chemistries are developed. One of the goals of the ExpoCast project is to assess the utility and reliability (Netzeva et al., 2005) of available existing physicochemical property models. This requires new experimental physicochemical data to be generated, but given the thousands of chemicals of potential interest, higher-throughput methods are attractive if they can be shown to be reliable.

Physicochemical properties have been found to be useful descriptors in predicting a wide range of properties (Kerns, 2001), including absorption (Camenisch, 2016), distribution(Peyret et al., 2010; Schmitt, 2008; Camenisch, 2016), clearance (Camenisch, 2016), permeability (Potts et al., 1992; Krämer, 1999), membrane (lipid bilayer) affinity (Yun et al., 2013; Pearce et al., 2017), plasma protein binding (Zhu et al., 2013; Ingle et al., 2016), in vitro assay concentration (Armitage et al., 2014; Fischer et al., 2017), and predictive ability of in vitro toxicokinetics (TK) assays (Wambaugh et al., 2015). Models derived from the use of physicochemical properties as well as molecular structure descriptors also enable the prediction of chemical functional use in products (e.g., emulsifiers and dyes) (Martin et al., 2010; Isaacs et al., 2016; Phillips et al., 2017). Thus, any resulting model's accuracy will be affected by uncertainty in the properties used, which subsequently impacts their utility for chemical risk assessment.

Physicochemical properties can be predicted from chemical structure via quantitative structure-property relationships (QSPRs) (Tong et al., 2004; Jaworska et al., 2005; Tropsha et al., 2007; Golbraikh et al., 2014; Dearden et al., 2009; Mansouri, 2013; Todeschini and Consonni, 2008). A QSPR expresses, in mathematical form, the quantitative relationship that may exist between the chemical structures of a series of chemicals and their measured properties. Many QSPRs are derived using machine learning algorithms which seek out statistically relevant correspondence between specific structural features and property values for a training set of chemicals (Tong et al., 2004; Jaworska et al., 2005; Tropsha et al., 2007; Golbraikh et al., 2014; Dearden et al., 2009; Mansouri, 2013; Todeschini and Consonni, 2008). Applicability domains (AD) are typically defined for QSPRs to facilitate reliable use. The AD for a QSPR is defined as the response and chemical structure space in which the model makes predictions with a given acceptable reliability. There are many different types of ADs that can be defined for QSPR models; for statistically based QSPR models relying on structural features, interpolation methods are often used (Tong et al., 2004; Jaworska et al., 2005; Tropsha et al., 2007; Golbraikh et al., 2014; Sahigara et al., 2012). Chemicals within the AD are associated with model-specific prediction uncertainty based on approximations to experimental measurements (Box, 1979). In addition, it is worth noting that prediction uncertainty is assumed to increase for chemicals determined to be outside of the AD or for models with unknown AD boundaries (Tong et al., 2004; Tropsha et al., 2007; Walker et al., 2003). Since predicted physicochemical properties are often used as inputs to derive QSAR models, whether they be for toxicity, environmental fate or toxicokinetic parameters, any prediction uncertainty will ultimately cascade to these model predictions as well (Walker et al., 2003). Unfortunately, for many chemicals of interest, the relevant physicochemical properties of interest have not been measured and/or are out of the predictive models' interpolation spaces (Arnot et al., 2012; USEPA, United States Environmental Protection Agency, Washington, 2017; Wambaugh et al., 2013).

Several QSPRs for physicochemical properties have been incorporated into software tools for ease of use. Some of these tools are open-source and free, such as OPEn (quantitative) structure-activity Relationship Application (OPERA) (Mansouri et al., 2018; Mansouri et al., 2016a), whilst others are proprietary but free, such as the Estimation Program Interface (EPI Suite) (USEPA, United States Environmental Protection Agency, Washington, 2017) and Online Chemical Database (OCHEM) (OCHEM, n.d.). There are also a number of property predictors that are proprietary and commercial, such as offered by Simulations Plus (n.d.) and Advanced Chemistry Development, Inc. (ACD/Labs) (Advanced Chemistry Development Inc, n.d.), or are embedded within software suites, such as ChemAxon (ChemAxon Ltd., n.d.) products that provide QSPR models. Some of these models are not transparent in terms of providing end-users with the information needed to assess their reliability, specifically details such as the AD, underlying training set and details of the model algorithms. For example, EPI Suite (USEPA, United States Environmental Protection Agency, Washington, 2017) does not provide an AD for any of its physicochemical QSPR models.

Here we describe a pilot study where we attempted to rapidly generate experimental data for 5 physicochemical properties of 200