



Contents lists available at ScienceDirect

Chemosphere

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

Predicting equilibrium vapour pressure isotope effects by using artificial neural networks or multi-linear regression – A quantitative structure property relationship approach

Julien Parinet^a, Maxime Julien^b, Pierrick Nun^b, Richard J. Robins^b, Gerald Remaud^b, Patrick Höhener^{a,*}

^a Aix-Marseille Université, Laboratoire Chimie Environnement, FRE 3416-CNRS, Marseille, France

^b Université de Nantes, Chimie et Interdisciplinarité: Synthèse, Analyse et Modélisation, UMR 6230-CNRS, Nantes, France

HIGHLIGHTS

- Equilibrium vapour pressure isotopes effects are predicted based on simple descriptors.
- An artificial neural network model was developed and validated.
- The neural network model was superior to a multi-linear regression model.
- The number of descriptors is critically discussed.

ARTICLE INFO

Article history:

Received 22 April 2014

Received in revised form 13 October 2014

Accepted 19 October 2014

Available online xxx

Handling Editor: I. Cousins

Keywords:

Equilibrium vapour pressure isotope effects

Quantitative structure–property relationship

Artificial neural networks

Multi-linear regression

ABSTRACT

We aim at predicting the effect of structure and isotopic substitutions on the equilibrium vapour pressure isotope effect of various organic compounds (alcohols, acids, alkanes, alkenes and aromatics) at intermediate temperatures. We attempt to explore quantitative structure property relationships by using artificial neural networks (ANN); the multi-layer perceptron (MLP) and compare the performances of it with multi-linear regression (MLR). These approaches are based on the relationship between the molecular structure (organic chain, polar functions, type of functions, type of isotope involved) of the organic compounds, and their equilibrium vapour pressure. A data set of 130 equilibrium vapour pressure isotope effects was used: 112 were used in the training set and the remaining 18 were used for the test/validation dataset. Two sets of descriptors were tested, a set with all the descriptors: *number of ¹²C, ¹³C, ¹⁶O, ¹⁸O, ¹H, ²H, OH functions, OD functions, C=O functions, Connolly Solvent Accessible Surface Area (CSA) and temperature* and a reduced set of descriptors. The dependent variable (the output) is the natural logarithm of the ratios of vapour pressures (*ln R*), expressed as light/heavy as in classical literature. Since the database is rather small, the leave-one-out procedure was used to validate both models. Considering higher determination coefficients and lower error values, it is concluded that the multi-layer perceptron provided better results compared to multi-linear regression. The stepwise regression procedure is a useful tool to reduce the number of descriptors. To our knowledge, a Quantitative Structure Property Relationship (QSPR) approach for isotopic studies is novel.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

Transport and transformation in the atmosphere are among the key processes that govern the distribution and fate of organic chemicals in the environment. Hence, an important feature of the partition of an organic compound in the environment is the quantitative description of how such a compound is distributed between the gas phase and other relevant (condensed) phases

(Schwarzenbach et al., 2003). Moreover, the vapour pressure of a compound is not only a measure of the maximum possible concentration of a compound in the gas phase at a given temperature, but it also provides important quantitative information on the attractive forces acting on the compound's molecules in the condensed phase. These properties are structure dependent (Godavathy et al., 2006; Gharagheizi et al., 2012). In the present study, we are concerned with the equilibrium vapour pressure isotope effect (EVPIE) of pairs of isotopologues, that is, molecules of the same species but with differing heavy isotope distributions. For two isotopologues with a heavy or light isotope, we define EVPIE as

* Corresponding author. Tel.: +33 413551034.

E-mail address: patrick.hohener@univ-amu.fr (P. Höhener).

$\ln R (= \ln P'/P)$, the ratio of the vapour pressures where P' and P are the vapour pressure of the light and heavy isotopologues, respectively. A positive $\ln R$ characterizes a normal isotope effect (the light isotopologue is more volatile), whereas a negative $\ln R$ points to an inverse isotope effect. As $\ln(1 + u) \approx u$ when u is small relative to 1, $-\ln R$ is approximately equal to the enrichment factor ε in environmental studies. The change in sign is needed to account for the different definitions light/heavy or heavy/light, which were used in classical and environmental literature, respectively. The EVPIE can have an important influence during the volatilization of organic compounds and pollutants (Wang and Huang, 2003; Höhener et al., 2008; Kuder et al., 2009; Thullner et al., 2012; Jeannotat and Hunkeler, 2012), making it of interest in environmental forensics (source appointment) and the monitoring of remediation. Studying vapour pressure isotope effects is a first step towards the ultimate goal of describing partitioning in more general terms, as e.g. applied for sorption to geosorbents (Höhener and Yu, 2012). Therefore, it is intended to predict EVPIE by a QSPR with easy to use inputs.

From a theoretical point of view, at a given temperature there are two opposite processes involved in establishing an EVPIE: evaporation and condensation of both isotopologues. These two processes coexist until equilibrium is reached. The first process can be described as a certain number of molecules thermally jostling in the condensed phase, a proportion of which will continuously acquire sufficient energy to overcome the forces of attraction to their neighbouring molecules and escape from the condensed phase. Meanwhile, in the vapour phase, there will be continuous collisions of some vapour molecules with the surface of the condensed phase. A fraction of the colliding molecules will have so little kinetic energy, or will dissipate their energy upon collision with the condensed surface, that rather than bounce back into the vapour phase, they will be combined into the condensed phase. Both processes are controlled primarily by molecule–molecule attractions in the condensed phase and are characterized by the amount of molecules in the vapour above the condensed phase. In other words, a compound will be liquid or a solid at a considered temperature if the forces between the molecules in the condensed phase are strong enough to overcome the tendency of the molecules to fly apart (Ivanciuc, 1999; Yao et al., 2001; Bagheri et al., 2012).

The prediction of EPVIEs is desirable for studies addressing the quantification of volatilization of liquids in the environment, such as water (Craig and Gordon, 1965), solvents (Jeannotat and Hunkeler, 2012), or petroleum hydrocarbons (Bouchard et al., 2008; Höhener et al., 2008). All these studies showed that the isotopic evolution during volatilization of organic liquids buried in soil is governed by two processes, the EVPIE and the diffusion effect. This diffusion effect is easy to predict since fractionation by diffusion is simply caused by differences in molecular mass and easily quantified by an equation originally given by Craig (see Craig and Gordon, 1965). For the EVPIE, no easy estimation method existed so far.

Measurement of an EVPIE for a pair of isotopologues was traditionally done by the manometric differential between a pure compound and its counterpart labelled at distinct positions with a heavy isotope. In some cases, measurements were also based on distillation (see Jancso and van Hook (1974) for details). Theoretical predictions of EVPIE were time consuming or very computational. Different models were proposed and summarized in the review by Jancso and van Hook (1974): Lindemann (1919); Friedmann (1962); Johns (1958) based on the Clausius–Clapeyron equation; Rabinovich (1962) who used a thermodynamic equation. Kiss et al. (1963) cited by Jancso and van Hook (1974) demonstrated empirically that, at a given temperature, EVPIE is correlated to the square root of the molecular weight of a compound. More

recently, Gharagheizi et al. (2012) used a Quantitative Structure Property Relationship (QSPR) molecular approach to predict the absolute vapour pressures on a very large dataset (1500 chemical compounds, mostly organic, but without considering isotopic substitution) by using an approach exploiting an artificial neural network (ANN). They obtained good results, with an absolute average relative deviation of the predicted values of about 7% ($r^2 = 0.990$) over a temperature range between 55 to around 3040 K. However, isotope effects are usually in the permil (‰) range and therefore cannot be predicted by this model.

Based on the work of Gharagheizi et al. (2012), we extended this approach to develop a QSPR model to predict EVPIE of 130 organic compounds (acids, alcohols, alkanes, alkenes, aromatics) with pairs of isotopic substitution exploiting a subset of the dataset from Jancso and Van Hook (1974). We excluded inorganic compounds and organic compounds with other functional groups. Easily accessible descriptors were first chosen (see Materials and Methods). However, with these descriptors, isomers like propan-1-ol and propan-2-ol are not differentiated, and therefore the Connolly Solvent Accessible Surface Area (CSA) was added as a further descriptor. This set of 11 descriptors (all-set) was used to feed a multi-layer perceptron (MLP) and a multi-linear regression (MLR) models. In order to test simplification, a reduced set (stepwise-set) with only six descriptors and without CSA was selected by a stepwise regression procedure made on all the dataset.

For each set with 11 or 6 descriptors, we compared the performance of MLP models with various numbers of neurons in the hidden-layers to find the optimal number of neurons. For the set of descriptors considered, we also compared the performance of the best MLP model to the MLR model. We further compared the performance of the best MLP model with all the descriptors to the best MLP model with the reduced set of descriptors. The same comparison was also done with the MLR models.

2. Materials and methods

2.1. Database

We extracted 130 vapour pressure isotope effects ($\ln R$) from 130 pairs of organic isotopologues (acids, alcohols, alkanes, alkenes, aromatics). The values of these EVPIE come from the review of Jancso and Van Hook (1974) and were measured at different temperatures. These organic compounds presented as pairs of isotopologues were treated for developing and validating the QSPR model. To the best of our knowledge, this is one of the largest databases available in the open literature of vapour pressure isotope effects.

Of the 130 $\ln R$ values, 112 were used in the training set and the remaining 18 were used for the test/validation dataset. To build this validation dataset, we chose eighteen values of $\ln R$ between all the four families (acids, alkanes, alkenes and alcohols) of isotopologues present in the training dataset at various temperatures.

2.2. Determination of the molecular descriptors

The choice of molecular descriptors is crucial. We worked within four constraints. Firstly, we were looking for easily-accessible molecular descriptors. Most models in the literature require data such as boiling temperatures or molar volumes, which are not available for each isotopologue. Secondly, the descriptors should characterize the two isotopologues of each pair. Thirdly, we need to identify descriptors that describe the link between structure and the desired property (EVPIE). Fourthly, none of the software available in the market was able to find descriptors to make the difference between two isotopologues: to distinguish a

Download English Version:

<https://daneshyari.com/en/article/6307753>

Download Persian Version:

<https://daneshyari.com/article/6307753>

[Daneshyari.com](https://daneshyari.com)