Supporting information

Mixture Touch: A web-platform for the evaluation of complex chemical mixtures

Yasuyuki Zushi†, Nobuyasu Hanari‡, Deedar Nabī§,ǂ, Bin-Le Lin†

†Research Institute of Science for Safety and Sustainability, National Institute of Advanced Industrial Science and Technology, National Institute of Advanced Industrial Science and Technology (AIST), 16-1 Onogawa, Tsukuba, Ibaraki 305-8569, Japan
‡Research Institute for Material and Chemical Measurement, National Metrology Institute of Japan, National Institute of Advanced Industrial Science and Technology (AIST) 1-1-1 Umezono, Tsukuba, Ibaraki 305-8563, Japan
§Institute of Environmental Sciences and Engineering, National University of Sciences and Technology, H-12, Islamabad, Pakistan
ǂCollege of Health Sciences, Jumeira University, Latifa Bint Hamdan Street, Dubai, United Arab Emirates

Table of contents

Section S-1. Mechanism of GC × GC property estimation ..........................................................S2
Section S-2. Algorithm of spectrum search ..............................................................................S3
References ..................................................................................................................................S3
Section S-1. Mechanism of GC × GC property estimation

The partitioning behavior of a compound between two different phases is expressed by the Gibbs free energy (\(\Delta G\)) of phase transfer for the target compound. The logarithm of the equilibrium constant between the phases (\(\log K_{xy}\)) is linearly correlated with the energy of phase transfer, known as the linear free energy relationship (LFER) shown in Eq. (1).

\[
\log K_{xy} = \frac{\Delta G}{RT} + c \ldots (1)
\]

where \(R\) is the ideal gas constant, and \(T\) is temperature of the system.

In the polyparameter LFER (pp-LFER) model, the \(\log K_{xy}\) is expressed by a linear combination of descriptors (\(E, S, A, B, L\)) of the chemical with system constants (\(e, s, a, b, l\)), as shown in Eq. (2).

\[
\log K_{xy} = eE + sS + aA + bB + ll + c' \ldots (2)
\]

where \(E\) is the descriptor for excess molar refraction, \(S\) is the descriptor for the molecular polarizability and polarity, \(A\) is the hydrogen-bond acidity, \(B\) is the hydrogen-bond basicity, and \(L\) is the logarithmic partition coefficient between air and hexadecane. There are several forms of the pp-LFER model, which are described in detail elsewhere (Endo & Goss, 2014; Panagopoulos et al., 2015).

Nabi et al. developed a two-parameter LFER, which uses descriptors extracted from the GC × GC retention times (Nabi & Arey, 2017; Nabi et al., 2014). For nonpolar chemicals, fewer dimensions are required to model \(\log K_{xy}\). Descriptors of the original pp-LFER were consolidated based on a singular value decomposition to two retention time descriptors (\(u_1\) and \(u_2\)), which were linked to partitioning coefficients in the first and second GC columns (Arey et al., 2005; Nabi et al., 2014). Eventually, properties related to partitioning can be expressed by the position of each non-polar chemical in the GC × GC as the following equations:

\[
\log K_{xy,i} = \lambda_1 u_{1,i} + \lambda_2 u_{2,i} + \lambda_3 \ldots (3)
\]

\[
u_{1,i} = \log L_{1,i} \ldots (4)
\]

\[
u_{2,i} = \log L_{2,i} - \beta_{orth} (\log L_{1,i}) \ldots (5)
\]

\[
\log L_{1,i} = \alpha_1 \cdot N_i + \alpha_2 \ldots (6)
\]

\[
\log L_{2,i} = \alpha'_1 \cdot N_i + \alpha'_2 + \log \left(\frac{t_{2,i}-a_1}{t_{2,i}-a_2}\right) \ldots (7)
\]
where $\lambda_{1-3}$ are the calibrated parameters from a multiple linear regression analysis in this study. The parameter $\beta_{orth}$ was determined by Gram–Schmidt orthogonalization. Parameters $\alpha_1$ and $\alpha_2$ were calibrated with the ideal carbon number ($N_i^*$) for selected non-polar chemicals using linear regression analysis. The $N_i^*$ for each chemical was calculated by linear interpolation of the retention times of alkanes (C$_{10}$–C$_{50}$) in the first dimension. Parameters $\alpha'_1$, $\alpha'_2$, and $\alpha_3$ were calibrated using $N_i^*$, $t_{2,i}$, and $t_{2,i}^*$ and non-linear regression analysis. The parameter $t_{2,i}$ is the retention time of the selected chemical in the second column, and $t_{2,i}^*$ is the RT2 of the ideal carbon where RT1 corresponds to that of a target compound. Further details for these calculations can be found elsewhere (Nabi & Arey, 2017; Nabi et al., 2014; Zushi et al., 2019). The Mixture Touch is equipped with an equation for 12 environmentally-relevant properties including logarithms of the partition coefficients of octanol–water ($\log K_{o-w}$) and air–water ($\log K_{a-w}$). The equation was calibrated with non-polar compounds, which were defined as structures containing carbon, hydrogen, and halogen atoms.

**Section S-2. Algorithm of spectrum search**

The Mixture Touch is equipped with a mass spectrum search function. A reference spectrum that shows a similar pattern to the target spectrum is picked from a database according to the degree of similarity. Practically, the similarity is calculated using the following equation:

$$\text{Similarity} = \left(\frac{1}{\sum Z_x^2} \otimes \frac{1}{\sum Z_y^2}\right) \times 1000 \ldots \ldots \ldots (8)$$

where $Z_x$ is the normalized intensity of the $m/z$ of the target spectrum, and $Z_y$ is its counterpart spectrum. $Z$ is the matrix of $Z_x$ and $Z_y$. The mass spectrum is generally evaluated using the NIST MS search program. According to the guidelines built into the NIST software, a value of 900 or greater is an excellent match, 800–900 is a good match, 700–800 is a fair match, and less than 600 is a very poor match. The above equation corresponds to the setting of the NIST software of ‘simple similarity’ with ‘default’ pre-search.

The URLs for external websites that also provide spectrum search functions with their own spectral libraries are available in the “Spectrum Search” tab. In its
current version, Mixture Touch is equipped with approximately 14,000 mass spectra provided by MS-DIAL website as open data (MS-DIAL).

References

Arey, J. S., Nelson, R. K., Xu, L., & Reddy, C. M. (2005). Using Comprehensive Two-Dimensional Gas Chromatography Retention Indices To Estimate Environmental Partitioning Properties for a Complete Set of Diesel Fuel Hydrocarbons. Anal. Chem., 77, 7172-7182.

Endo, S., & Goss, K.-U. (2014). Applications of Polyparameter Linear Free Energy Relationships in Environmental Chemistry. Environ. Sci. Technol., 48, 12477-12491.

MS-DIAL. FAQ. [http://prime.psc.riken.jp/Metabolomics_Software/MS-DIAL/index3.html](http://prime.psc.riken.jp/Metabolomics_Software/MS-DIAL/index3.html) Accessed 28.2.2020.

Nabi, D., & Arey, J. S. (2017). Predicting Partitioning and Diffusion Properties of Nonpolar Chemicals in Biotic Media and Passive Sampler Phases by GC × GC. Environ. Sci. Technol., 51, 3001-3011.

Nabi, D., Gros, J., Dimitriou-Christidis, P., & Arey, J. S. (2014). Mapping Environmental Partitioning Properties of Nonpolar Complex Mixtures by Use of GC × GC. Environ. Sci. Technol., 48, 6814-6826.

Panagopoulos, D., Jahnke, A., Kierkegaard, A., & MacLeod, M. (2015). Organic Carbon/Water and Dissolved Organic Carbon/Water Partitioning of Cyclic Volatile Methylsiloxanes: Measurements and Polyparameter Linear Free Energy Relationships. Environ. Sci. Technol., 49, 12161-12168.

Zushi, Y., Yamatori, Y., Nagata, J., & Nabi, D. (2019). Comprehensive two-dimensional gas-chromatography-based property estimation to assess the fate and behavior of complex mixtures: A case study of vehicle engine oil. Sci. Total Environ., 669, 739-745.