Towards methods for facile estimation of enthalpies of vaporization of neutral organic chemicals using the principles of linear free energy relationship

By

Manahil Rafique

(Reg# 00000320353)

A thesis submitted in partial fulfillment of requirements for the degree of

Master of Science

In

Environmental Science

Institute of Environmental Sciences and Engineering (IESE)

School of Civil and Environmental Engineering (SCEE)

National University of Sciences and Technology (NUST)

Sector H-12, Islamabad, Pakistan

2021

CERTIFICATE

It is certified that the contents and forms of the thesis entitled '**Towards methods for facile estimation of enthalpies of vaporization of neutral organic chemicals using the principles of linear free energy relationship'** submitted by **Manahil Rafique** has been found satisfactory for the requirements of MS degree.

Supervisor:

Dr. Deedar Nabi: ___________________

Associate Professor IESE, SCEE, NUST

GEC Member: _____________________

Dr. Zeehan Sheikh

Associate Professor

IESE, SCEE, NUST

GEC Member: ______________________

Dr. Hira Amjad

Assistant Professor

IESE, SCEE, NUST

THESIS ACCEPTANCE CERTIFICATE

It is certified that the copy of MS thesis entitled by **Ms. Manahil Rafique** Registration No. 00000320353 of IESE (SCEE) has been vetted by the undersigned, found complete in all aspects as per NUST Statutes/Regulations, is free of plagiarism, errors, and mistakes and is accepted as partial fulfillment for the award of MS degree. It is further certified that necessary amendments as pointed out by GEC members have also been incorporated in the said thesis.

Signature with stamp: ___________________

Name of the Supervisor: Dr. Deedar Nabi

Date: _________________

Signature of HoD with stamp: _________________

Date: ______________________

Countersign by

Signature (Dean/Principal): __________________

Date: __________________

DECLARATION

I certify that research work titled "*Towards methods for facile estimation of enthalpies of vaporization of neutral organic chemicals using the principles of linear free energy relationship"* is my work. This work has not been presented elsewhere for assessment. Where material has been obtained from other sources, it has been properly acknowledged/referred.

Manahil Rafique

00000320353

“This thesis is dedicated to my Mentor Dr. Deedar Nabi, my affectionate parents, and my uncle.”

Acknowledgments

I was not able to accomplish this research work without the guidance of Allah Almighty, the most beneficent and merciful. I would like to express my utmost gratitude to my Supervisor, Dr. Deedar Nabi for his understanding, wisdom, patience, and continuous motivation. I would also like to express gratitude to my GEC members, Dr. Hira Amjad and Dr. Zeshan Sheikh. Their affection, guidance, and expertise generously helped me during my research. I am also thankful to my family, friends, and research group fellows to encourage me in stressful situations and not let me give up. Lastly, I would like to thank my parents for their unconditional love, support, and encouragement.

Table of Contents

CERTIFICATE 2

List of Abbreviations 9

ABSTRACT 10

INTRODUCTION 11

1.1 Background 11

1.2. Problem statement 13

1.3. Hypothesis 14

1.4. Scope of the study 14

1.5. Significance of the study 14

Chapter 2 15

LITERATURE REVIEW 15

METHODOLOGY 19

3.1. Data acquisition 19

3.2. Calibration and evaluation of model 19

3.2.1. 2-parameter partitioning model 19

3.2.2. GC x GC Model 21

3.3. Statistical analysis 22

Chapter 4 24

RESULTS AND DISCUSSIONS 24

4.1. Justification of two-parameter model 24

4.2. Two-parameter LFER model 27

4.2.1 2-parameter model developed on ASM and Experimental log Kow and log Kaw values 27

4.2.2 2-parameter model developed on Experimental $log\ Kow$ and $log\ Kaw$ values 28

4.2.3 2-parameter model trained on the log Kow and log Kaw values estimated from EPI-Suite 28

4.2.4 2-parameter model developed using interaction terms 29

4.3. Model Validation 32

4.3.1 2p-PM validation 32

4.3.2 GC x GC model validation 35

CONCLUSIONS AND RECOMMENDATIONS 38

5.1 Integration of partitioning model into EPI Suite^{TM} 38

5.2 Limitations and outlooks 39

REFERENCES 40

Supporting Information..47

List of Abbreviations

$\Delta_{\mathrm{vap}}\mathrm{H}_{\mathrm{m}}^{0}$	Standard molar enthalpy of vaporization
$\Delta_{\mathrm{solv}}\mathrm{H}_{\mathrm{m}}^{0}$	Standard molar enthalpy of solvation
2p-PM	Two parameter partitioning model
LFER	Linear Free Energy Relationship
POPs	Persistent Organic Pollutants
GC	Gas Chromatography
LFER	Linear Free Energy Relationship
OP-LFER	One Parameter-Linear Free Energy Relationship
ASM	Abraham Solvation Model
QSAR	Quantitative Structure-Activity Relationship
PP-LFER	Poly Parameter-Linear Free Energy Relationship
Kow	Octanol-Water Partition Coefficient
Kaw	Air-Water Partition Coefficient
PCA	Principal Component Analysis
AIC	Akaike Information Criterion
LOOCV	Leave One Out cross validation

ABSTRACT

Recently, research interest is revolving in understanding the temperature dependence of environmental partitioning properties governing the fate, behavior, and transport of organic pollutants, which may be attributed to the global warming phenomenon. The environmental partitioning property data are generally measured at room temperature (20-25˚C) which requires the temperature correction for the hot climatic regions. Enthalpy is an important thermodynamic parameter required to correct partitioning property data for temperatures differences. In this thesis, I developed an easy and parsimonious 2-parameter partitioning model (2p-PM) to predict standard molar enthalpies of vaporization. Unlike previous models such as the widely used Abraham Solvation Model (ASM), my 2p-PM — based on correlation of enthalpy of vaporization with a linear combination of two partition coefficients of octanol-water and air-water systems — is computationally fast and parametrically parsimonious with almost similar predictive performance as observed for the ASM. This new model can be integrated in the US EPA's EPI-Suite Software. In the second part of my thesis, I developed a GC×GC model which is based on the retention time information of non-polar chemicals on the comprehensive two-dimensional gas chromatography (GC×GC). It can be applied to complex environmental mixtures such as polyhalogenated flame retardants, paint additives, and plasticizers. Taken together, this study provides the means and methods to understand the temperature dependence of vaporization of chemicals in different climatic scenarios.

Keywords

Standard molar enthalpy of vaporization, linear free energy relationship, Enthalpy of phase change, Abraham solvation model, quantitative structure vaporization enthalpy relationships, Trouton's rule, entropy, EPI-Suite™.

Chapter 1

INTRODUCTION

1.1 Background

Organic chemicals' equilibrium partitioning properties stimulate their transportation and circulation among various phases of the environment such as air, soil, water, etc. (Schwarzenbach, R. P., Gschwend, P. M., & Imboden, 2002). Fate models typically work with equilibrium partitioning properties for evaluating their behavior along with examining the effects of hydrophobic pollutants on the environment. They work as input parameters, which are defined as to

$$P_{xy,i} = \left\{\frac{C_{x,i}}{C_{y,i}}\right\} \, equillibrium \qquad (1)$$

$where,$

$P_{xy,i} = Partition\ coefficient\ amid\ x\ and\ y\ phase$

$C_{x,i}, C_{y,i} =$

$At\ partitioning\ equilibrium, contaminent\ i\ concentration\ in\ two\ phases$

Partitioning of a compound between two media is influenced by factors like solubility, pH of the system, temperature etc. (Bahadur et al., 1997)(Chaurasia, 2017)(Chiou et al., 1977).

If we measure partition coefficients as a function of temperature so it requires lengthy and accurate experimentation. The dependence of partition coefficient on temperature is an important factor but most often it is not considered while developing environmental models based on partitioning parameters.(Bahadur et al., 1997).

Jianguo Wu et al. used global climate models for estimating variations in $K_{soil-air}$ values for POPs under warming climatic conditions. It was found that Climate warming not only strongly influences the temporal and spatial variability of contaminants but also their behavior, partitioning properties, and persistence etc (Wu, 2020). Recent studies showed that Global climate change lead to many processes and interactions that have strong potential to affect the physiochemical properties and their fate and transport in the environment is also altered because partitioning of chemicals is sensitive to it (Gouin et al., 2013)(Wu, 2020). Therefore, it is necessary to estimate

or model the organic chemicals' fate and transport in the environment based on the temperature-dependent environmental-related partitioning coefficients (Macleod et al., 2007)**.**
In gas-phase systems, partitioning of chemicals is highly temperature-dependent so their respective enthalpies of the phase transfer (enthalpy of vaporization) must be known.

Standard molar enthalpy of vaporization is the measure of energy that is required by one mole of the substance, at constant temperature and pressure, to undergo a transition from liquid to gas phase. The usual unit is kilojoules per mole (kJ/mol) (Helmenstine, Anne Marie, Ph.D (2020).

An improved Watson equation for the enthalpy of vaporization as a function of temperature is given in the following equation (Coker, 2014).

$$\Delta H_v = A(1 - TT_c)n \qquad (2)$$

Where
ΔH_v = enthalpy of vaporization (kJ/mol)
$A, T_c \; and \; n$ = regression coefficients for a chemical compound
T = temperature (K)

The logarithm of the partition coefficient of compound i is linearly related to the corresponding enthalpy by following free energy relation.

$$lnK_i = -\frac{\Delta G}{RT} + consti \qquad (3)$$

As we know

$$\Delta G_i = \Delta H_i - T\Delta S_i \qquad (4)$$

By using above equation 3, it is possible to estimate the the equilibrium partition coefficient of compound i at temperature T if the enthalpy of phase change is known. (Goss & Schwarzenbach, 1999).

Above equation 4 shows a quantitative linear relationship between enthalpy and entropy of a partitioning system but this relationship can be implied only as long as chemicals are dominated by the same type of interactions. (Goss & Schwarzenbach, 1999). Abraham and coworkers (Abraham & Acree, 2012) also highlighted a predictive method for estimating K_w (gas to water partition coefficient) values at temperature range of 273-373K via equations proposed by Plyasunov and Shock. This method is based on using the thermodynamic input quantities, *log* K_w (298), $\Delta H_{w,}$ and ΔC_{pw} for the estimation of K_w (gas to water partition coefficient) values. But the problem while considering this approach is that it requires all the three thermodynamic properties for predicting K_w (gas to water partition coefficient) values.

Chromatographic techniques can also be used as a tool for the estimation of various physical and biological properties. Various studies explore the potential of gas chromatography (Ellison, 2005) (Gobble & Chickos, 2015) (Chickos et al., 1995) to estimate the enthalpies of vaporization of neutral organic compounds. In recent studies, two-dimensional gas chromatography emerged as a better chromatographic technique for the risk assessment of chemicals (Barden & McGregor, 2017). According to Nabi et al., many diffusion-related (*logK*) and environmental partitioning properties of nonpolar complex organic compounds were favorably predicted using LFERs based on two solute parameters i.e., $u_{1,i}$ and $u_{2,i}$ which were extracted from the first- and second-dimensions retention time of the analytes on GC×GC chromatogram. A set of 79 nonpolar model chemicals was first used to theoretically calibrate the GC×GC model Eq. 1 for 32 properties and then we validated our model equation 1 via set of 52 nonpolar chemicals already analyzed on the two-dimensional gas chromatography GC×GC instrument.

$$logK = \lambda_3 + \lambda_2 u_{2,i} + \lambda_1 u_{1,i} \qquad (5)$$

Where $\lambda_3, \lambda_2, \lambda_1$ are constants and specific to each partitioning system. This GC×GC model is a powerful tool that allows the estimates of properties to apply directly on the nonpolar compounds detected in environmental mixtures (Naseem et al., 2021), (Nabi & Arey, 2017).

1.2. Problem statement

It is difficult to measure the enthalpy values for chemicals experimentally due to the unavailability of optimum conditions and environmental losses (material losses, energy losses, changes in temperature, pressure, etc.) during experiments. Also, the existing computational models to find enthalpy values have limited applicability because some of these are parameter intensive and need laborious experimental work to find out the descriptors, some of these do not have enough databases of required independent parameters, and some of these are too difficult for interpretation of non-technical users. These practical constraints invited us to develop simple and parsimonious models to reliably find the enthalpy values of chemicals.

1.3. Hypothesis

2p-PM based on the linear combination of *logKow* and *logKaw* (octanol-water and air-water portioning coefficients respectively) and GCxGC model based on descriptors $u_{1,i}$ and $u_{2,i}$ can predict enthalpies of vaporization of neutral organic chemicals with the accuracy nearly equal to Abraham Solvation Model.

By keeping in view the above problem statement, we designed our study based on the following objectives.

- To develop a statistically robust, rigorous, and parsimonious partitioning model that would be computationally inexpensive and simple to apply
- To investigate different types of inter-molecular descriptors dictating the enthalpies of the phase transfer for neutral organic chemicals.
- To develop a robust GC x GC model that would have potential to predict enthalpies of phase change for non-polar complex mixtures.

1.4. Scope of the study

The research work was divided into two phases.

- In the first phase, 2p-partitioning model has been developed to determine variability in enthalpy of phase change for neutral organic chemicals.
- In the second phase, the model has been tested for certain criteria of internal and external validity to check its robustness.

1.5. Significance of the study

- Our newly developed two-parameter models (both partitioning and GC x GC) can be used by various researchers in chemical laboratories or industries to estimate enthalpies of phase change rapidly and accurately at low cost.
- Estimated enthalpy data can be incorporated into the partitioning model to make necessary temperature corrections due to climate change.

Chapter 2

LITERATURE REVIEW

For last many years, researchers are putting their considerable efforts to accurately measure $\Delta_{vap}H_m^0$ values. The experimental value of this thermodynamic property for any chemical is not easy to measure accurately due to certain limitations like human errors, environmental losses, and unavailability of optimum conditions. At present experimental measurements from researchers are unable to keep pace with the discovery of new chemical compounds. Moreover, a decrease in spending from the private sector for research and development has also caused reduction in development of experimental values.

So, to cater this problem, we seek help from predictive methods or techniques. For estimating $\Delta_{vap}H_m^0$ values of pure chemicals, already available predictive methods can be categorized as those based on:

(i) Trouton's rule and other associated methods (Fishtine, 1963a)(Fishtine, 1963b) (Wadsö et al., 1966)(Zhao, Li, et al., 1999)(Zhao, Ni, et al., 1999).

(ii) Correlations of vapor pressure with $\Delta_{vap}H_m^0$ and information about critical constants (Clausius-Clapeyron equation) (ANTOINE & C., 1888)(Wagner, 1973)(Walton, 1989)

(iii) Law of corresponding states (PITZER et al., 1993)(Wang & Shi, 1990)(Morgan & Kobayashi, 1994).

(iv) Estimation methods involving empirical mathematical equations which highlight relationship between vaporization enthalpy and other measured physical properties (Giacalone, n.d.)(Chen, 1965)(Vetere, 1979)(Vetere, 1995)(Liu, 2001)(Bowden & Jones, 1948)(Wright, 1960)(An et al., 1995).

(v) QSAR methods by using descriptors derived from structural considerations (Ivanciuc et al., 2001)(Arjmand & Shafiei, 2018)(Abooali & Sobati, 2014)(Gharagheizi, 2012).

(vi) Estimation method such as group additivity, group contribution, and fragmentation methods used for assigning the numerical values to the functional group or atom arrangement present in the molecule (Naef & Acree, 2017)(Abdi et al., 2018)(Rebas et al., 2016)(Gharagheizi et al., 2011).

(vii) Models based on quantum mechanics (Kaminski et al., 2017) (Flôres et al., 2016).

(viii) Models based on Abraham solute descriptors. (Churchill et al., 2019) (Shanmugam et al., 2021)(Abraham et al., 2021)(Tirumala et al., 2020)(Shanmugam et al., 2021).

Hence, there is an emerging trend since many years to develop fast and authentic estimation methods for enthalpies of phase change.

There are certain limitations associated with each predictive method. For example, Group additivity and group contribution methods can be applied only to those molecules for which all required group values are available. (Churchill et al., 2019).

Models based on QSAR (quantitative structure activity relationship) method estimated $\Delta_{vap}H_m^0$ values at normal boiling points with good correlations. But this method required the molecular descriptors and normal boiling data of chemicals.

Another predictive method based on Solvation free energy calculations makes the use of electrostatic contribution COSMO solvation model for the electrostatic contribution, and perturbation theory for the hard-core molecules' cavity term. But this method was limited to the class of organic compounds comprising only the selective kinds of atoms (H, C, F, Cl, N, O atoms). Moreover, model also required the experimental data of liquid molar volumes which is often not readily available or requires difficult alternatives. (Lin et al., 2004).

Recently, Solomonov and coworkers published a remarkable work in which they reported the method to estimate both standard molar enthalpies of vaporization and sublimation based on the difference in enthalpies of solution and solvent.

$$\Delta_{sub}H_m^0 = \Delta_{soln}H_m^0 - \Delta_{solv}H_m^0 \quad (6)$$

$$\Delta_{vap}H_m^0 = \Delta_{soln}H_m^0 - \Delta_{solv}H_m^0 \quad (7)$$

But this method also has limitations associated with it as it requires a wide range of solvents and applies to limited number of chemicals. (Solomonov et al., 2004).

Churchill and coworkers, in their study developed poly-parameter LFER model equations comprising Abraham solute descriptors to accurately estimate $\Delta_{vap}H_m^0$ for 703 chemicals. It was found that Abraham model correlations also provide reasonably accurate estimates of $\Delta_{vap}H_m^0$ so they consider the same modeled equations for $\Delta_{vap}H_m^0$ estimation that has already been used to estimate $\Delta_{solv}H_m^0$.

Churchill model showed that enthalpy of phase change depends upon different types of intermolecular interactions that are believed to be present in solution.

$$\Delta_{vap}H_m^0\ (kJmol^{-1}) = 5.938(0.313) - 7.667(0.456)E + 9.983(0.876)S + 15.483(1.200)A + 1.694(0.558)B + 9.608(0.067)L - 1.541(0.618)S.S + 43.483(1.964)A.B \quad (8)$$

$$(where\ n = 703, SD = 2.55, R2 = 0.979, F = 4710.9)$$

$$\Delta_{vap}H_m^0\ (kJmol^{-1}) = -3.246(0.412) + 5.114(0.515)E + 19.635(0.955)S + 20.131(1.355)A + 1.266(0.629)B + 34.388(0.271)V - 2.487(0.698)S.S + 42.350(2.215)A.B \quad (9)$$

$$(Where\ n = 703, SD = 2.87, R2 = 0.974, F = 3687.6)$$

Where, E, S shows polarizability/di-polarity and molar refraction of solute, A, B represents acidity and basicity of the hydrogen bonds present in solute, V represents McGowan characteristic volume of a solute when molecules are stationary(cm^3 /mole)/10 and L is the logarithm of the solute's dimensionless gas-to-hexadecane partition coefficient at 298 K (Churchill et al., 2019)(Naseem et al., 2021)(Holley et al., 2011).To yield better mathematical correlations they introduced two additional interaction terms named as A.B and S.S to account for any compound-compound interactions that could be present in pure organic compounds.

We can see that modeled equations (8) (9) have good explanatory power but a careful point-by-point comparison of calculated versus experimental $\Delta_{vap}H_m^0$ values reveal that correlation overpredicted enthalpy values for 28 alkylamines ,16 alkanediols and 1 alkanetriol. Therefore, new model correlations were then developed after excluding these chemicals from the list comprising of 658 chemicals.

$$\Delta_{solv}H_m^0\ (kJmol^{-1}) = 6.192(0.243) - 7.688(0.361)E + 10.222(0.684)S + 3.068(1.366)A + 1.341(0.506)B + 9.517(0.052)L - 1.038(0.483)S.S + 81.336(3.314)A.B \quad (10)$$

$$(Where\ n = 658, SD = 1.95, R2 = 0.986, F = 6759.9)$$

$$\Delta_{solv}H_m^0\ (kJmol^{-1}) = -2.960(0.356) + 4.688(0.452)E + 20.076(0.863)S + 8.803(1.711)A + 0.328(0.633)B + 34.145(0.236)V - 1.861(0.606)S.S + 77.495(4.147)A.B \quad (11)$$

$$(With\ N = 658, SD = 2.44, R2 = 0.979, F = 4286.1)$$

A good R^2 and reduced SD values show that back calculations were poor for above-mentioned compounds.

Further in their study Churchill et al.,2019, introduced 3 indicator variables to their already modeled equations to include these chemicals into Abraham model correlations.

$$\Delta_{solv}H_m^0\ (kJmol^{-1}) = 6.100(0.257) - 7.363(0.380)E + 9.733(0.733)S + 4.025(1.351)A + 2.123(0.521)B + 9.537(0.055)L - 1.180(0.515)S.S + 77.871(3.233)A.B - 5.781(0.441)I_{amine} - 14.783(1.235)I_{non-\alpha,\omega-diol} - 17.873(1.431)I_{\alpha,\omega-diol} \quad (12)$$

$$(With\ N = 703, SD = 2.09,\ R2 = 0.986, F = 4925.6)$$

$$\Delta_{solv}H_m^0\ (kJmol^{-1}) = -3.008(0.368) + 5.226(0.456)E + 18.422(0.892)S + 8.978(1.661)A + 1.363(0.637)B + 34.141(0.242)V - 2.045(0.631)S.S + 75.728(3.952)A.B - 4.888(0.543)I_{amine} - 13.297(1.510)I_{non-\alpha,\omega-diol} - 17.619(1.748)I_{\alpha,\omega-diol} \quad (13)$$

$$(With\ N = 703, SD = 2.54, R^2 = 0.979, F = 3274.7)$$

It has been seen that all Abraham modeled correlation equations proposed by the authors have good explanatory power, but they require computationally expensive Abraham experimental input parameters. Currently, only for <8000 organic compounds experimental data of Abraham descriptors is available. Moreover, above correlations require different indicator variables to include chemicals that showed deviations. So, all these problems call for the model that could precisely estimate $\Delta_{vap}H_m^0$ of compounds without using Abraham descriptors and involving any indicator variable.

Chapter 3

METHODOLOGY

In this chapter, detail about data acquisition, materials and methods is described which was used for the investigation purpose of our study. Reliable computational software like R-studio, XLSTAT were used to perform data analysis.

3.1. Data acquisition

Standard molar enthalpy of vaporization $\Delta_{vap}H_m^0$ experimental values involving 703 chemicals were taken from the referenced paper (Churchill et al., 2019). List of all chemicals along with the values of their respective ASDs is mentioned in the table S1 of supporting information. To avoid over-representation, multiple values reported for a single chemical were averaged using an arithmetic mean. All the inorganic values were omitted from the data sets. We included all the chemicals (28 alkylamines,16 alkanediols and 1 alkanetriol) into the data list which had been eliminated into the previous reference study because our proposed 2-parameter LFER can account for enthalpy estimates for all chemicals providing their K_{ow} and K_{aw} values are known. The values for Abraham solvation descriptors, SMILES codes, and CAS numbers of chemicals were taken from the freely available database UFZ LSER. The estimated values of K_{ow} and K_{aw} were acquired from EPI Suite ^{TM} 4.1 – KOWWIN v1.68, Henry Win v 3.20, (US-EPA, 2018) respectively. The experimental values of $logK_{ow}$ were available for only 355 chemicals out of 703 chemicals. We used the ASM model (Poole et al., 2013) to calculate the $logK_{ow}$ values for the remaining 348 chemicals. Similarly, the experimental values of $logK_{aw}$ were available for 398 chemicals the remaining values were predicted by using ASM (Poole et al., 2013). We used the values of $logK_{aw}$ and $logK_{ow}$ estimated from ASM in place of unavailable experimental values.

3.2. Calibration and evaluation of model

3.2.1. 2-parameter partitioning model

For calibration and evaluation of 2-parameter model, we will use K_{ow} and K_{aw} values estimated via ASM equations(Poole et al., 2013) instead of the experimental values because for most of the chemicals experimental values were not available (for 355 chemicals exp. K_{ow} values and for 398 chemicals K_{aw} values). We observed that in comparison to the other estimation approaches, ASM estimated K_{ow}, K_{aw} values were more accurately.

For further evaluation of accuracies, we compared both ASM predicted, and EPI-suite estimated values of $logK_{ow}$ and $logK_{aw}$ with experimental data reported in reference(Churchill et al., 2019)

A careful point-by-point comparison of the ASM and EPI suite estimated values for the same experimental data (n=355) reveals that ASM equation for $logK_{aw}$ performed much better (*RMSE*=0.24) than KOWWIN v1.68 (*RMSE*=0.25). Similarly, ASM equation for $logK_{aw}$ was better (*RMSE*=0.38) than Henry Win v3.20(*RMSE*=0.45) when the experimental data size was n=398. So, when experimental values are available, we would prefer them over ASM predicted values, which in turn would be considered as more accurate or preferable over the EPI-Suite estimated values of $logK_{ow}$ and $logK_{aw}$.

The 2-parameter model is then evaluated by putting experimentally determined and EPI Suite™ estimated values of K_{ow} and K_{aw}. For the final data set, both experimental as well as estimated values of K_{ow} and K_{aw} are attached as an additional file in table S2 of supporting information.

We used ASM predicted K_{ow} and K_{aw} values only for a rigorous evaluation of the model. Once this evaluation is done, our 2p-LFER model does not need ASDs any longer as K_{ow} and K_{aw} properties in the model are not dependent only on ASDs for their values. $\Delta_{vap}H_m^0$ values of the chemicals for which values of ASDs are not known-can be estimated by using logK_{ow} and logK_{aw} in the 2p-LFER model. Further, these parameters can be either measure in a laboratory, or can be found in literature, or can be predicted reliably by using estimation approaches like 4.1 – KOWWIN v1.68, Henry Win v3.20 (US-EPA, 2018). Conversely, it has been seen that laboratory measurements of *Kow* and *Kaw* are easy than laboratory measurements of ASDs.

In the end, we tested the fitting of our 2p-LFER model to experimental data of $\Delta_{vap}H_m^0$, $logK_{ow}$ and $logK_{aw}$ (n=300) (R^2=0.83). We also tested its fitting on data containing experimental values of $\Delta_{vap}H_m^0$ and EPI Suite™ estimated values of $logK_{ow}$ and $logK_{aw}$ (R^2=0.83). Statistics of these trainings were then compared to the statistics of model when it was trained on data involving experimental $\Delta_{vap}H_m^0$ values and ASM estimated $logK_{ow}$ and $logK_{aw}$ values (R^2=0.94).

3.2.2. GC x GC Model

The GC × GC Model was calibrated and evaluated by using the data from the previous study (Nabi et al., 2014). A set comprised of 79 nonpolar chemicals from different chemical families was used for the calibration of the model (as shown in table 1).

Table 1. Set of 79 nonpolar chemicals used for the calibration of the GC × GC model

Chemical	**Chemical**	**Chemical**	**Chemical**
nonane	carbon tetrachloride	benzene, propyl-	fluorene
decane	1,1,2-trichloroethane	benzene, butyl-	phenanthrene
undecane	hexachloroethane	benzene, pentyl-	pyrene
dodecane	γ-HCH	benzene, octyl-	benz[a]anthracene
methylcyclopentane	1,3-butadiene, 1,1,2,3,4,4-hexachloro-	benzene, decyl-	chrysene
cyclooctane	1,3-cyclopentadiene, 1,2,3,4,5,5-hexachloro-	fluorobenzene	PCB 28
cyclododecane	Enflurane	1,3-difluorobenzene	PCB 52
cyclohexadiene	1-bromobutane	1,4-difluorobenzene	PCB 101
1,5,9-cyclododecatriene	1-bromooctane	1,3,5-trifluorobenzene	PCB 118
3-methylcyclohexene	dibromomethane	1,2,3,5-tetrafluorobenzene	PCB 138
cyclonona-1,2-diene	tribromomethane	benzene, 1,3-dichloro-	PCB 153
fluoromethane	hexabromoethane	benzene, 1,4-dichloro-	PCB 180
1-fluorobutane	diiodomethane	benzene, 1,2-dichloro-	p,p'-DDE
1-fluoropentane	1,2-diiodethane	benzene, 1,2,4-trichloro-	
1-fluorononane	1-iodohexane	benzene, hexachloro-	
tetrafluoromethane	Iodononane	bromobenzene	
sulfur hexafluoride	1-iodobutane	1,4-dibromobenzene	
1-chlorobutane	Benzene	1,3-dibromobenzene	
1-chlorooctane	Toluene	1,3,5-tribromobenzene	
1,4-diiodobenzene	acenaphthylene	1,2,3,5-tetrabromobenzene	
naphthalene	Acenaphthene	iodobenzene	
naphthalene, 1-methyl-	Dibenzofuran	1,3-diiodobenzene	

The singular value decomposition (SVD) analysis was done on 6 ASDs (E, S, A, B, L, V) of 79 chemicals to further evaluate the representativeness of the calibration dataset. The first dimensions explain the 99% variability according to SVD analysis. In the next step, the solute parameters u_1

and u_2 for the calibration set of 79 chemicals were acquired by the transformation of the gas-stationary phase partition coefficients for the first and second dimensions of the GC × GC. The ASM equations for the relevant stationary phases from the literature were used to estimate the values of the gas-stationary phase partition coefficient for these 79 nonpolar chemicals (Churchill et al., 2019) . Then, to develop two-parameter GC x GC, we used u_1 and u_2 as independent variables and $\Delta_{vap}H_m^0$ as dependent variable and performed MLR analysis.

In the final step, the above fitted GC × GC model was independently validated by using earlier published (Nabi et al., 2014) u_1 and u_2 values for 52 nonpolar chemicals. For this set the solute parameters u_1 and u_2 were acquired by transforming the retention time of first and second dimension of nonpolar analytes measured by GC × GC instrument (Nabi et al., 2014), (Nabi & Arey, 2017). The calibration set and the validation set were different in a sense that the values of input parameters u_1 and u_2 for validation set were experimentally determined by the analysis of these chemicals on GC × GC instrument while the values of u_1 and u_2 were determined theoretically for the calibration set.

For some nonpolar chemicals, experimental $\Delta_{vap}H_m^0$ values were not available in the validation and calibration set of the GC × GC model. So, for such chemicals we used the ASMs' estimated values. After the training and validation, GC × GC model no longer needs the experimental values of ASDs. In contrast to ASMs, the GC × GC models now can be used to predict $\Delta_{vap}H_m^0$ values for complex non-polar mixtures. The only thing that users need would be the values of u_1 and u_2 of a chemical to estimate its $\Delta_{vap}H_m^0$ values. The u_1 and u_2 values for the nonpolar chemicals can be easily determined by analyzing them on the GC × GC instrument. To develop GC × GC model, both validation and training datasets contain only nonpolar chemicals which have representatives of many chemical families such as benzene, *n*-alkanes, cycloalkanes, halogenated alkanes, cycloalkenes, halogenated alkenes, n-alkylbenzenes, halogenated benzenes, polychlorinated naphthalenes (PCNs), polychlorinated biphenyls (PCBs), polycyclic aromatic hydrocarbons (PAHs), polybrominated diphenyl ethers (PBDEs), and organochlorine pesticides.

3.3. Statistical analysis

In my study, I performed statistical analyses like multiple linear regression, Principal Component Analysis (PCA), and cross-validation by using statistical software R-studio (version - 3.5.3) (R (3.5.3), n.d.)(Dexter, 2014) and XLSTAT (2018). Contribution of a variable in the model was considered statistically significant if the computed t-value of the variable coefficient is less than or equal to the critical t-values reported at the significance level (p-value <0.05) for a given degree of freedom. The Akaike Information Criterion was employed for the selection of the ideal number of variables in the model. AIC penalizes the model upon adding new variables that do not impart sufficient information to the model. Hence, a model with minimum AIC value was selected. Analysis of correlation was also performed to check any overlapping information brought by different descriptors.

Different regression diagnostics were applied to the models e.g studentized residuals, Cook's distance, and hat values, in order to determine its domain of applicability and to identify the influential values in the training datasets. Moreover, bootstrapping technique was used for estimating the standard errors of beta-coefficients (fitting co-efficient) in my model. Some cross-validation techniques such as K-nearest neighbors, K-fold (*n* = 10), repeated K-fold (*n = 10, repeat = 3, repeat=10*), leave-one-out and bootstrapping (*n* = 1000) to evaluate the robustness. To identify the contribution of each variable in the principal component, PCA test was used.

Chapter 4

RESULTS AND DISCUSSIONS

4.1. Justification of two-parameter model

I proposed a hypothesis that enthalpy of phase change of neutral organic and organometallic compounds can be estimated adequately by using two parameters K_{ow} and K_{aw}. To test this hypothesis, I examined the information content that each descriptor in Abraham solvation parameter model contains. Abraham solvation parameter model equations from the literature shows that for neutral organic compounds, five dimensions of the information *[E, S, A, B, L]* and *[E, S, A, B, V]* are good enough to successfully explain about 98% and 97% of the variability in the enthalpy data respectively. However, PCA on the 703×7 matrix of ASDs *[E, S, A, B, V, L] of* the training set shows that the first two of the total 7 dimensions express about 71.7% of the total information. Rest of the information 28.3% was depicted by the remaining dimensions. It has been seen that the first dimension was primarily comprised of a linear combination of L, S, B and V with small influence of A and E. while the second dimension was comprised of mainly S, V and the remaining dimensions were comprised of minor contributions from other descriptors. Results show the distribution of variance among six parameters. It also originates a need for the development of a parsimonious model in replacement of such parameter intensive model. But what could be appropriate descriptors in the parsimonious model that would express the maximum information corresponding to all these dimensions of PCA?

Criteria for selection of model descriptors

The following considerations should be taken into account while selecting suitable descriptors:

i. They should be accessible easily.
ii. They can be either easily measurable in laboratory or has a wide experimental database.
iii. They could sufficiently account for changes in free energy during any phase change.

In fig 1, it can be observed that octanol-water and air-water partition coefficients are the two unique parameters that have been qualified for the above criteria. To prove this claim, we reviewed the information distribution resulting from the PCA 703×9 matrix *[E, S, A, B, V, L,* ΔHm*, log*K_{ow} and *logK$_{aw}$]* in principal components. It has been seen that maximum variability in enthalpy data has been partitioned between the first two dimensions (76%) of PCA and so my two partition coefficients (*logK$_{ow}$* and *logK$_{aw}$).* Only these two descriptors alone contain much of the information encoded into all ASDs.

The *ggheat* map of the correlation matrix shows the correlation of our two descriptors, *logK$_{ow}$* and *logK$_{aw}$* with $\Delta_{vap}H_m^0$, indicating that they contain similar information about intermolecular interactions, encoded in ASDs. Taken together, it has been seen that, *logK$_{ow}$* and *logK$_{aw}$* show a strong correlation (r= 0.3, r=-0.43 respectively) with $\Delta_{solv}H_m^0$. Hence, supporting results from correlation encourages us to prefer the suitability of *K$_{ow}$* and *K$_{aw}$* properties as these properties have easy and quick estimation approaches with a wide experimental database compared to ASDs.

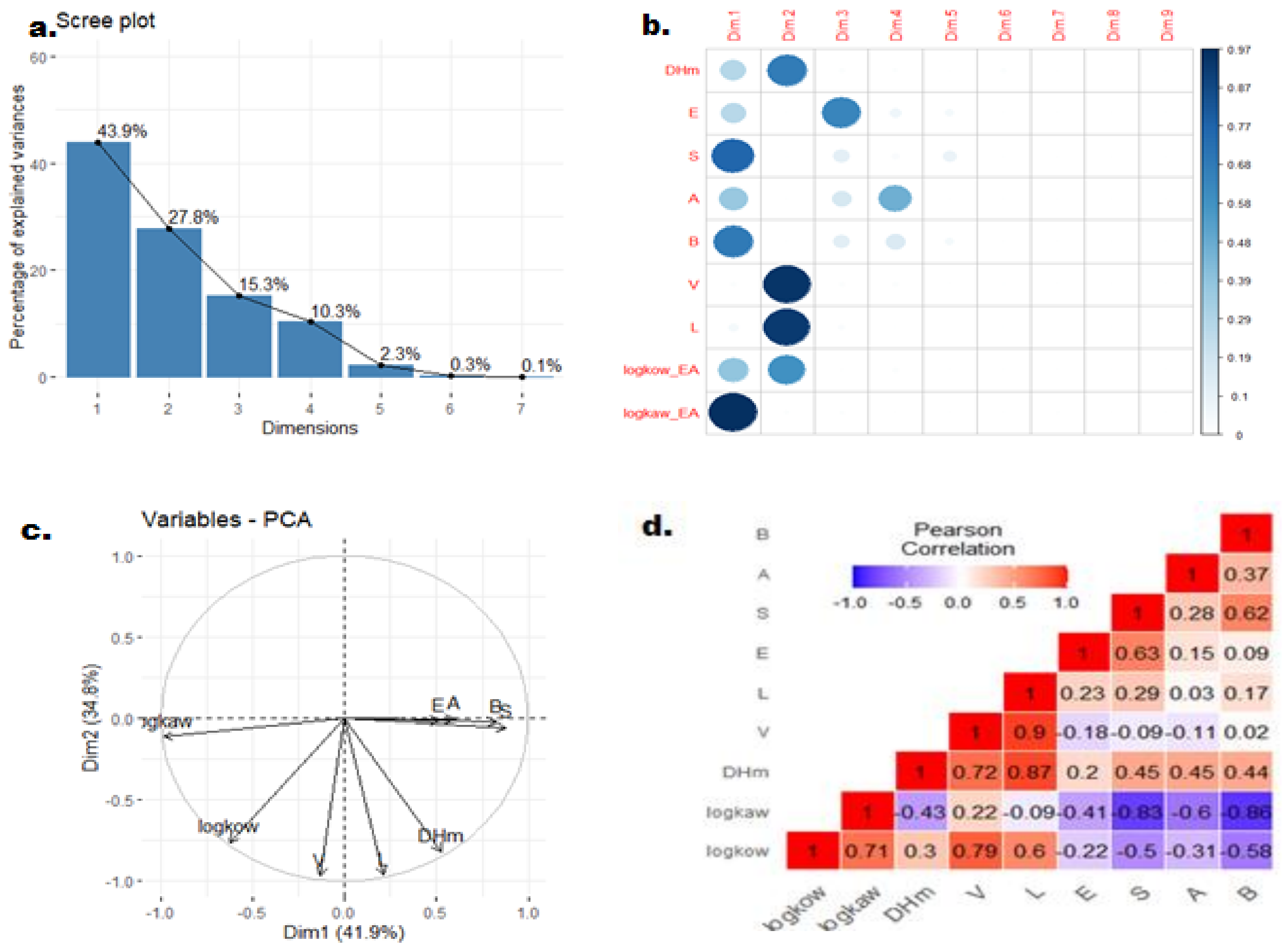


Fig.1 Dimensionality analysis of my proposed 2-pm. First three boards display the results obtained from principal component analysis performed on the 703x 7 matrix of Abraham solute descriptors. **a.** scree plot of eigenvalues (i.e., how much information each principal component retains). **b.** correlogram representing the distribution of Cos2 (quality of representation) into 8 dimensions. size and color intensity of circle is proportional to the quality of presentation of information by variable **c.** correlation circle representing the quality of representation and relationship between variables in first two dimensions. The length of the arrowed line from origin to the circumference is proportional to the quality of representation of the variable. Angles between these arrowed lines measure of the degree of correlation among descriptors. *V* and *L* are almost orthogonal to the *[E, S, A, B]* but show a positive correlation with each other. **d.** gg heat correlation map (correlogram obtained after Pearson correlation analysis of 703x 9 matrix). Colors red and blue indicate positive and negative correlations respectively among different pairs of variables.

4.2. Two-parameter LFER model

4.2.1 2-parameter model developed on ASM and Experimental log K_{ow} and log K_{aw} values

My two parameter Linear model based on the descriptors, *logK_{ow}* and *logK_{aw}* successfully explained the variability in $\Delta_{solv}H_m^0$ data with R^2=0.91.

$$\Delta_{solv}H_m^0(kJmol^{-1}) = 5.460(\pm 0.567) + 10.876(\pm 0.141)logK_{ow} - 9.418(\pm 0.115)logK_{aw}$$
(14)

(n=703, R^2=0.9134, Adj R^2 =0.91, RMSE= 5.224)

Here the n denotes the number of experimental observations of $\Delta_{vap}H_m^0$, R^2 and Adj R^2 represents the coefficient of correlation and adjusted coefficient of correlation respectively and RMSE stands for root mean square error. To train our model equation, we estimate the values of K_{ow} and K_{aw} from respective ASM equations due to the scarcity of experimental data (Poole et al., 2013).

First, I put experimental values of K_{ow} and K_{aw} in the model equation, tested its performance then put the ASM predicted values and tested its performance. We found that experimentally determined K_{ow} and K_{aw} values were in good agreement with the $\Delta_{vap}H_m^0$ experimental values than ASM predicted *log*K_{ow} and *log*K_{aw}. These statistics suggested that *log*K_{ow} and *log*K_{aw} experimental values exhibit more accuracy than ASM predicted values.

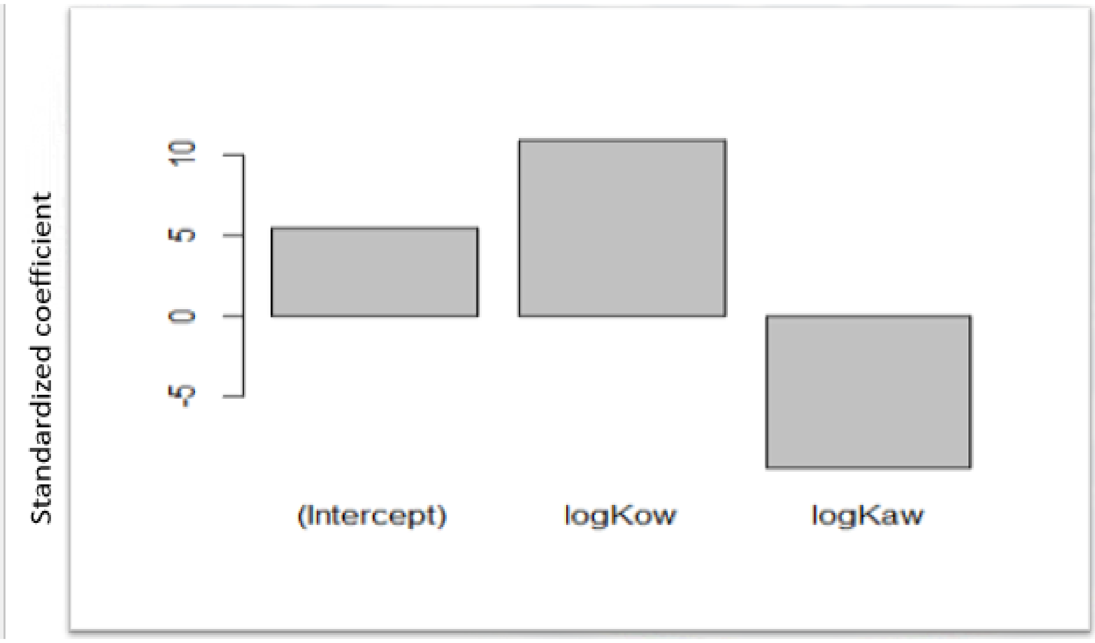


Figure 2: Bar-plot representing the contribution of each partition coefficient

4.2.2 2-parameter model developed on Experimental *log* K_{ow} and *log* K_{aw} values

When the 2-parameter model was trained solely on experimentally determined $\Delta_{solv}H_m^0$, *logKow* and *logKaw* values, it resulted in the regression coefficient of R^2 = 0.84.

$$\Delta_{solv}H_m^0(kJmol^{-1}) = 8.473(\pm1.082) + 9.677(\pm0.304)logK_{ow} - 9.067(\pm0.239)logK_{aw}$$
(15)

$$(\text{n} = 300, R^2 = 0.84, \text{Adj } R^2 = 0.84, \text{RMSE} = 4.088)$$

Where n represents the number of experimental values of $\Delta_{solv}H_m^0$, *logKow* and *logKaw*. R^2, Adj $R^2$and RMSE represents the correlation coefficient, adjusted correlation coefficient and root mean square error respectively.

4.2.3 2-parameter model trained on the log K_{ow} and log K_{aw} values estimated from EPI-Suite.

When 2-parameter model was trained solely on EPI-Suite estimated *logKow* and *logKaw* values, it was successful to explain about 84% variability in my enthalpy data.

$\Delta_{solv}H_m^0(kJmol^{-1}) = 6.545(\pm 0.788) + 11.43(\pm 0.211)logK_{ow} - 9.032(\pm 0.161)logK_{aw}$ (16)

$$(\text{n} = 703, R^2 = 0.84, \text{Adj } R^2 = 0.84, \text{RMSE} = 7.175)$$

Where n represents the number of experimental values of $\Delta_{solv}H_m^0$, *logK*$_{ow}$ and *logK*$_{aw}$. R^2, Adj $R^2$and RMSE represents the coefficient of determination, adjusted coefficient of determination and root mean square error respectively.

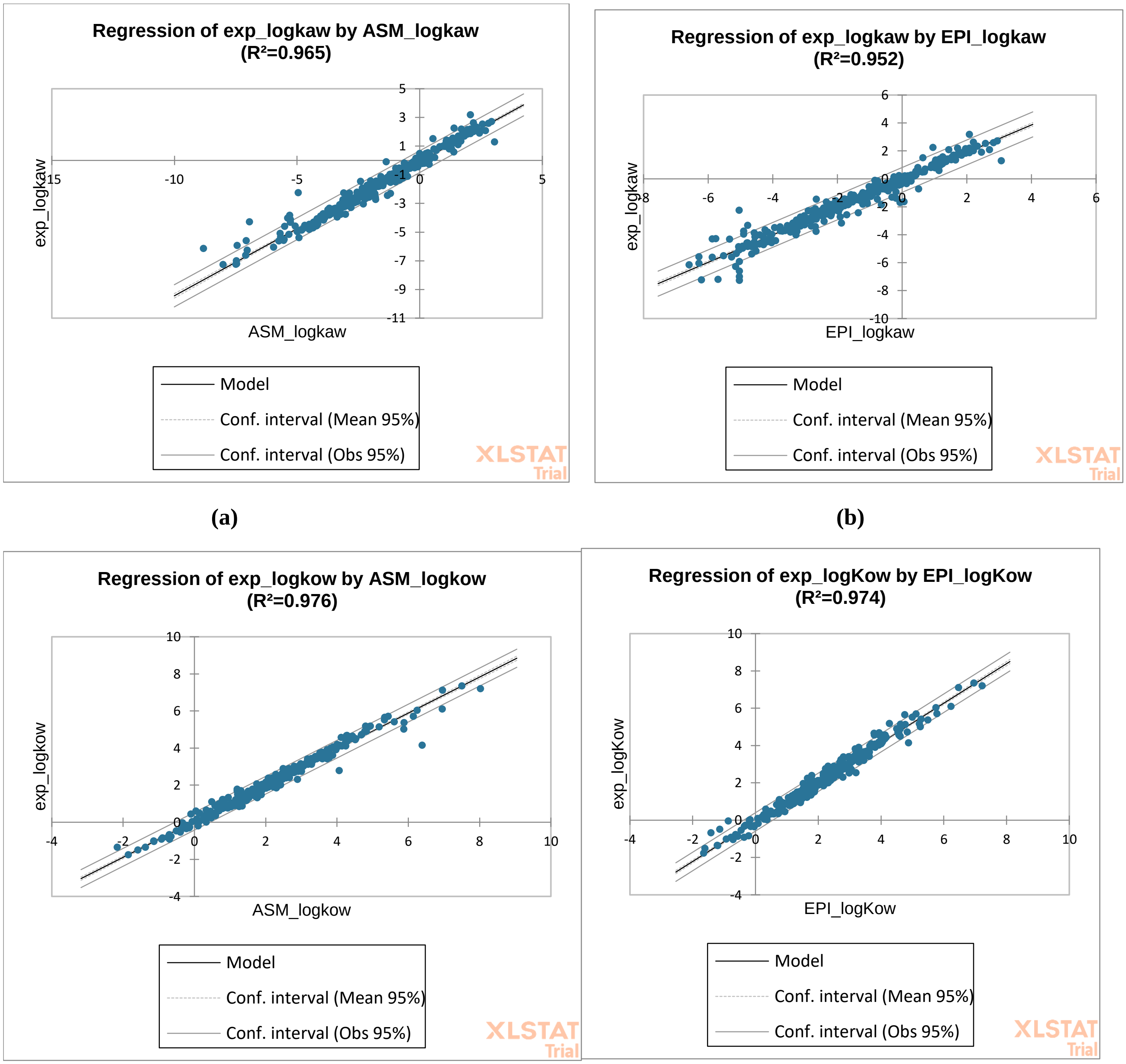
4.2.4 2-parameter model developed using interaction terms

Churchill et al., introduced interactions terms *A.B* and *S.S* in their Abraham model correlation equations for compound-compound interactions that could be present in pure organic compounds and produced significantly better correlation results. Addition of interaction terms in our already developed modeled equation 4 yielded a slightly better mathematical correlation.

$$\Delta_{solv}H_m^0(kJmol^{-1}) = 4.677(\pm 0.561) + 10.934\ (\pm 0.137)logK_{ow} - 9.958\ (\pm 0.136)\ logK_{aw} + 0.250\ (\pm 0.036)\ logK_{ow} * logK_{aw} \qquad (17)$$

$$(\text{n} = 703, R^2 = 0.9188, \text{Adj } R^2 = 0.9185, \text{RMSE} = 5.061)$$

From the above equation we can see that addition of interaction term in the already developed model equation for the 703 chemical data set didn't cause any significant increase in R^2 value (from 0.914 to 0.918) but eq 6 is recommended over eq 5 as it deals with the compound-compound interaction already been discussed in the referenced Churchill model.



(a) **(b)**

(c) **(d)**

Fig 3: (a) Shows that experimental logkaw is plotted against ASM logkaw. (b) Shows that experimental logkaw is plotted against estimated logkaw. (c) Shows that experimental logkow is plotted against ASM logkow. (d) Shows that experimental logkow is plotted against estimated logkow. Upper and lower lines around the regression line(solid line in the middle) bound the 95% confidence interval.

4.2.4 One parameter LFER model

In the end I will check the contribution of the explanatory power of descriptors K_{ow} alone to explain the variability in $\Delta_{vap}H_m^0$.

$$\Delta_{solv}H_m^0(kJmol^{-1}) = 0 + 12.11\ (\pm\ 0.2897) logK_{ow} \qquad (18)$$

$$(\text{n} = 703, R^2 = 0.7134, \text{Adj } R^2 = 0.713, \text{RMSE} = 27.99)$$

$$\Delta_{solv}H_m^0(kJmol^{-1}) = 0 - 10.2128\ (\pm\ 0.63)\ logK_{aw} \qquad (19)$$

$$(\text{n} = 703, R^2 = 0.2722, \text{Adj } R^2 = 0.2712, \text{RMSE} = 44.59)$$

Results have shown that out of a total 91.4% variance, 71% is explained by K_{ow} alone. In comparison, K_{aw} depicts a small proportion of variability in $\Delta_{vap}H_m^0$. Reason behind the unequal depiction of information by two variables, K_{ow} and $K_{aw,}$ is that the former accounts for all the specific and non-specific interactions that are believed to be present in a compound while the later accounts for only hydrogen-bond interactions.

4.2.5 GC x GC model

My newly developed GC × GC model (Eq.20) is successful to explain the variability in the $\Delta_{vap}H_m^0$ data of nonpolar organic chemicals with good correlation coefficient $R^2 = 0.988$. Here, only the experimental $\Delta_{vap}H_m^0$ data of 79 nonpolar chemicals was used to create training set. ASM estimated values were preferred when the experimental values were unavailable.

$$\Delta_{solv}H_m^0(kJmol^{-1}) = 11.980(\pm 0.718) + 16.397(\pm 0.239)u_1 - 10.277(\pm 1.983)u_2 \qquad (20)$$

$$(n = 79, \quad R^2 = 0.988, \quad Adj. R^2 = 0.987\ , \quad RMSE = 2.393)$$

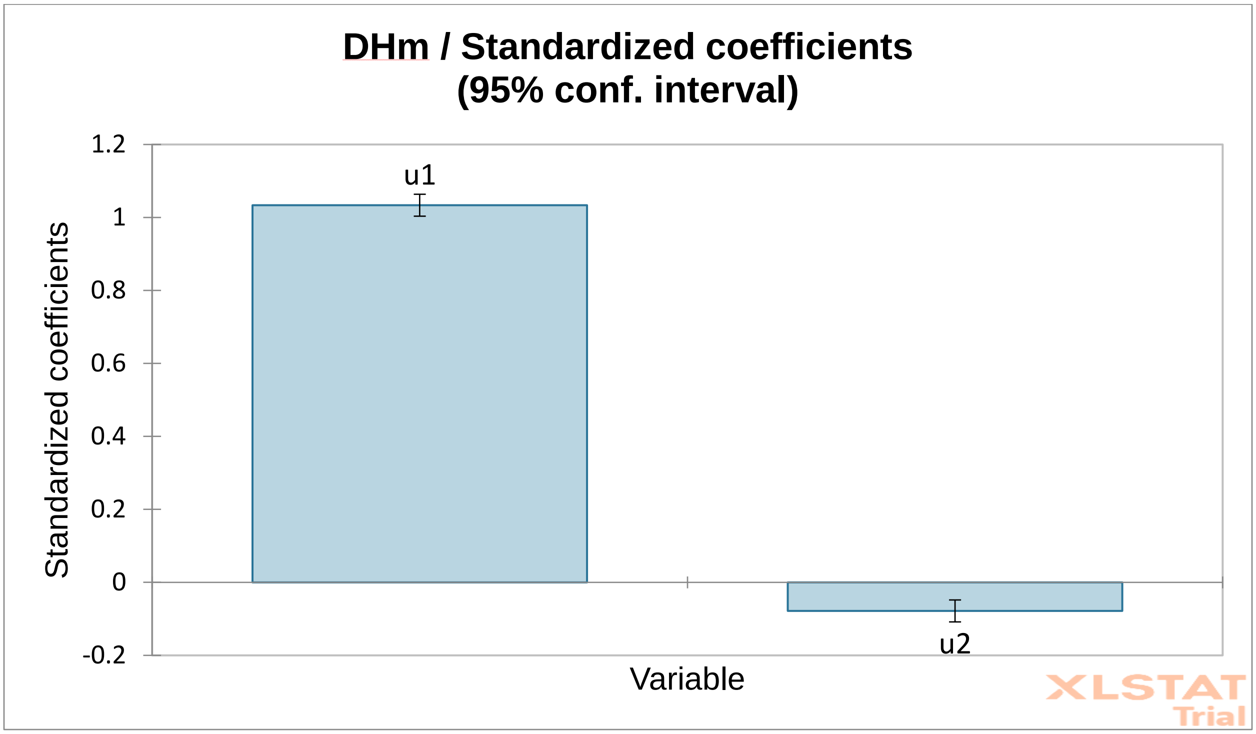


Figure 4: Bar-plot representing the contribution of each partition coefficient

4.3. Model Validation

4.3.1 2p-PM validation

I tested this model (eq 4) for certain criteria of external and internal validity. For internal validation, I performed four independent cross-validation tests (K-fold, repeated K-fold, Leave-one-out, and Bootstrap cross-validations) with their results (see Table S1) indicating that the developed model is statistically robust. It fulfills the criteria for internal validity and can be used for predictive purposes.

Table S1: Cross-validation of 2-parameter model (equation 4) to evaluate model robustness.

Indicators	**Hold-out approach (splitting of data, test vs train)**		**LOOCV**	**K - fold CV**	**Repeated K - fold CV**		**CV by Bootstrapping**		
	KNN(test)	KNN(train)			3 times	10times	N= 100	N=500	N=1000
R2	0.9091	0.9144	0.912	0.912	0.913	0.912	0.914	0.911	0.911
RMSE	5.25	5.22	5.25	5.144	5.15	5.14	5.217	5.25	5.26
Adj.R2	0.9087	0.9141	------	------	-------	-------	---	----	----
MAE	3.551	3.327	3.39	3.39	3.39	3.39	3.404	3.41	3.42

For external validation, I will randomly split our whole data set into a training set (n=564) and test set (n=139) through the Hold-out approach. Model is first trained on a training set comprising 564 chemicals resulting in equation 5.

$$\Delta_{solv}H_m^0(kJmol^{-1}) = 5.5165(\pm 0.63) + 10.875(\pm 0.156)logK_{ow} - 9.360(\pm 0.12)logK_{aw}$$
(21)

(n=564, R^2=0.9144, Adj R^2 =0.9141, RMSE= 5.223)

It has been seen that equation 21 has regression statistics similar to the equation 14. Moreover, cross-validation statistics of both equations were also found to be the same. For both training and validation sets, I compared the predicted $\Delta_{vap}H_m^0$ values from equation 5 with experimental values. As equation 4 is being trained on full data set, so it would be more appropriate and highly recommended to the users to prefer equation 4 trained on full data set (n = 703) than equation 5.

Table S2: Cross-validation statistics of 2p-PM (equation 5) by using 4 independent tests.

Indicators	**LOOCV**	**K - fold CV**	**Repeated K - fold CV**		**CV by Bootstrapping**		
			3 times	10times	N= 100	N=500	N=1000
R2	0.912	0.915	0.915	0.915	0.914	0.912	0.9131
RMSE	5.26	5.144	5.15	5.14	5.20	5.27	5.24
Adj.R2	------	------	-------	-------			
MAE	3.35	3.36	3.36	3.35	3.36	3.38	3.38

It has been seen that model equation trained solely on available experimental K_{ow} and K_{aw} values yielded better mathematical correlations than trained on ASM estimated values of $logK_{ow}$ and $logK_{aw}$. However, the mathematical correlation was found to be worsened when the model equation was trained on the $logK_{ow}$ and $logK_{aw}$ values estimated from EPI-Suite (KOWWIN v1.69 and Henry win v3.21). So, for appropriate estimation, we would recommend preferring equation 4 above all the trained equations as it is being trained on the large and accurate dataset. But if, for some chemicals, experimental $logK_{ow}$ and $logK_{aw}$ values and ASM based $logK_{ow}$ and $logK_{aw}$ values are not available so in this case $logK_{ow}$ and $logK_{aw}$ values estimated from EPI-Suite would be a good alternative.

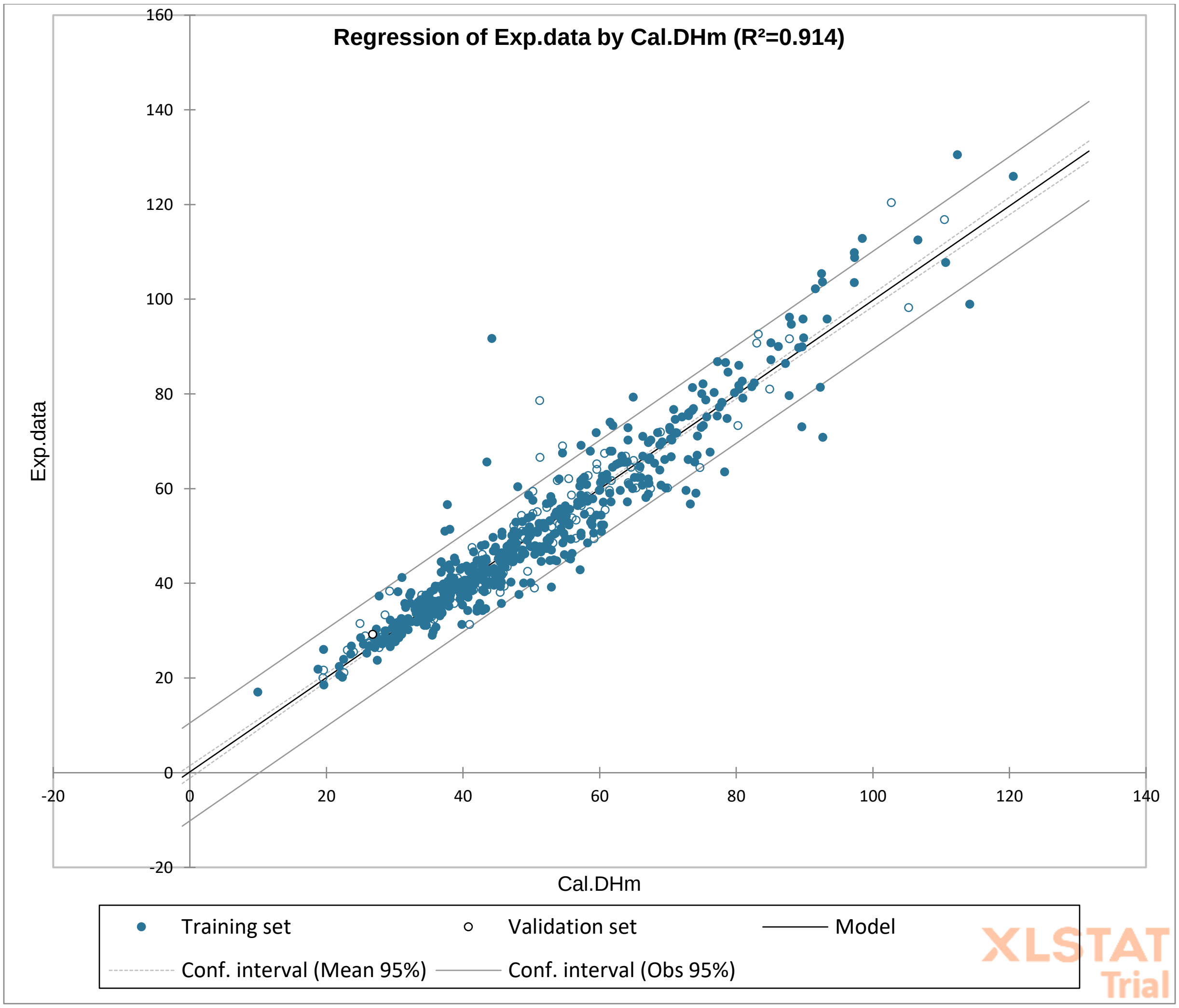


Figure 5: Experimental VS calculated $\Delta_{solv}H_m^0$. Linear regression plot for Two - Parameter Partitioning Model (eq 4) showing training set and validation set. Here upper and lower dotted lines bound 95% CI around the regression line (dotted middle line).

4.3.2 GC x GC model validation

The performance of our GC×GC model was evaluated by comparing its estimated values with the ASMs and PMs predicted $\Delta_{vap}H_m^0$ values. The RMSE value was (3.2153) for the GC×GC model while the RMSE values of PMs was (9.397) for the same model set. The cross-validation was done by using the bootstrap and leave-one-out techniques to further evaluate the model robustness.

I would use the following independent approach to validate the GC×GC models. The input parameters u_1 and $u_2$were acquired by the analysis of 70 nonpolar chemicals on GC×GC instrument from an earlier study (Nabi et al., 2014). The values of u_1 and u_2 for this dataset were incorporated into already developed GC×GC model equation. The calculated values of $\Delta_{vap}H_m^0$by this method compared appropriately with $\Delta_{vap}H_m^0$ values estimated by ASMs and PM.

Table S2: Cross-validation statistics of GC × GC model by using independent tests

Indicators	**Hold-out approach (splitting of data, test vs train)**		**LOOCV**	**K - fold CV**	**Repeated K - fold CV**		**Bootstrapping method N=500**	**Independent set approach**
	KNN(test)	KNN(train)			3 times	10times		N=70
R2	0.984	0.989	0.9865	0.912	0.98	0.99	0.98	-----
RMSE	3.122	2.149	2.44	5.144	2.40	2.28	2.46	3.215
MAE	2.01	1.40	1.66	3.39	1.66	1.67	1.72	-----

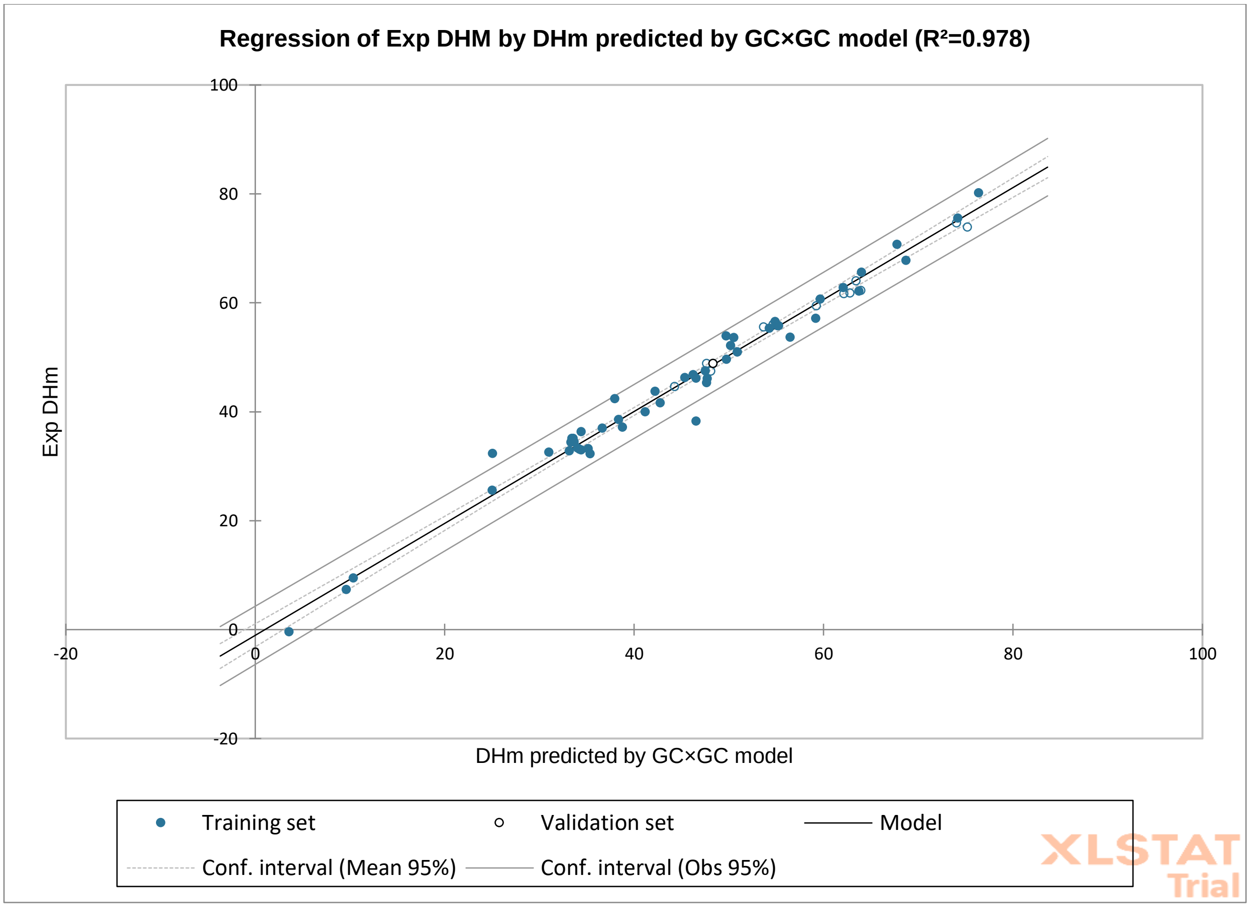


Figure 6: Experimental VS calculated $\Delta_{solv}H_m^0$ via GC × GC model. Linear regression plot for GC × GC model showing training set and validation set. Here, upper and lower dotted lines bound 95% CI around the regression line (dotted middle line).

Chapter 5

CONCLUSIONS AND RECOMMENDATIONS

I have been able to develop a poly-parameter linear regression equation, based on two descriptors K_{ow} and K_{aw} from EPI Suite™, which correlates the $\Delta_{vap}H_m^0$ data with squared correlation coefficient of approximately 91%. Above derived correlation in eq 4 is comparable to, if not better than, other method that make use of complicated and expensive techniques for estimating $\Delta_{vap}H_m^0$. My model performed better than Churchill model, in a sense, that it can predict $\Delta_{vap}H_m^0$ values of chemicals belonging to any class of neutral organic compounds for which Abraham descriptors are not known and without using indicator variables. Above all, Churchill correlation model is more accurate as it gives us approximately 98% accurate estimates of $\Delta_{vap}H_m^0$ but at cost of expensive Abraham descriptors which are currently available only for 8000 or so organic compounds. From here my model takes the lead as it can predict enthalpy change values without using above descriptors with an accuracy nearly equal to Churchill model. We believe that this study, to some extent, overcome the limitations of already developed models and paved a pathway leading to accurate, facile and, rapid risk assessment of organic species via enthalpy data.

5.1 Integration of partitioning model into EPI Suite^{TM}

Estimation program interface (EPI) Suite^{TM} is the collection of many modules for the estimation of environmental fate and physical/chemical properties of a large number of compounds. It was developed by the Environmental Protection Agency (EPA) and Syracuse Research Corporation(Balakrishnan et al., 2020). I inputted the estimated values of $\log K_{ow}$ and $\log K_{aw}$ given by EPI Suite^{TM} software (KOWWIN v1.69 and Henrywin v3.21) into the PMs for enthalpies of phase changes. The values predicted by using this approach were compared with the experimental values of $\Delta_{vap}H_m^0$ obtained from the literature (Churchill et al., 2010). Though the statistics (RMSE ranged from 0.3988 to 1.280) were not as good as for experimental and ASM predicted $\log K_{ow}$ and $\log K_{aw}$ values (RMSE ranged from 0.2134 to 0.5817) but still, these are good enough for integration into EPISuite^{TM} for the easy and quick predictions of $\Delta_{vap}H_m^0$values for multiple neutral organic compounds.

5.2 Limitations and outlooks

A disadvantage of my 2-parameter model is that it can make predictions only for neutral organic compounds. It is not suitable for ionized species which show distinct partitioning behavior than neutral compounds.(Bouwer, 1997). Moreover, it is applicable only under neutral conditions. However, by adding descriptors like pK_{α} according to a given pH of the system of interest it is possible to account for the partitioning behavior of ionized species. (Franco & Trapp, 2008). Introducing the descriptors of ionizability to the model and its evaluation would extend the scope of my study that would be discuss in future studies.

Due to unavailability of experimental $logK_{ow}$ and $logK_{aw,}$ I trained our model by taking the estimated values of $logK_{ow}$ and $logK_{aw}$ via ASM equations.(Poole et al., 2013). Although, ASM equations give us accurate predictions of $logK_{ow}$ and $logK_{aw}$ values, but I believe that if our 2-parameter model has been trained on experimental values its predicting power is expected to be improved. But if I train the model on pure experimental data, inflated errors around regression coefficients are expected.

REFERENCES

1. Abdi, S., Movagharnejad, K., & Ghasemitabar, H. (2018). Estimation of the enthalpy of vaporization at normal boiling temperature of organic compounds by a new group contribution method. *Fluid Phase Equilibria*, *473*, 166–174. https://doi.org/10.1016/j.fluid.2018.06.006
2. Abooali, D., & Sobati, M. A. (2014). Novel method for prediction of normal boiling point and enthalpy of vaporization at normal boiling point of pure refrigerants: A QSPR approach. *International Journal of Refrigeration*, *40*, 282–293. https://doi.org/10.1016/j.ijrefrig.2013.12.007
3. Abraham, M. H., & Acree, W. E. (2012). The hydrogen bond properties of water from 273 K to 573 K; Equations for the prediction of gas-water partition coefficients. *Physical Chemistry Chemical Physics*, *14*(20), 7433–7440. https://doi.org/10.1039/c2cp40542c
4. Abraham, M. H., Acree, W. E., & Liu, X. (2021). Descriptors for High-Energy Nitro Compounds; Estimation of Thermodynamic, Physicochemical and Environmental Properties. *Propellants, Explosives, Pyrotechnics*, *46*(2), 267–279. https://doi.org/10.1002/prep.202000117
5. An, W. F., Li, J. Y., Cheng, W. W., & Gao, J. D. (1995). Estimation of heat of vaporization for pure compounds with a residual function method. *The Chemical Engineering Journal and The Biochemical Engineering Journal*, *59*(2), 101–109. https://doi.org/10.1016/0923-0467(94)02917-2
6. ANTOINE, & C., M. (1888). Nouvelle Relation Entre les Tensions et les Temperatures. *C. r. Held Seanc. Acad. Sci. Paris*, *107*, 681–684.
7. Arjmand, F., & Shafiei, F. (2018). Prediction of the Normal Boiling Points and Enthalpy of Vaporizations of Alcohols and Phenols Using Topological Indices. *Journal of Structural Chemistry*, *59*(3), 748–754. https://doi.org/10.1134/S0022476618030393
8. Bahadur, N. P., Shiu, W. Y., Boocock, D. G. B., & Mackay, D. (1997). Temperature dependence of octanol-water partition coefficient for selected chlorobenzenes. *Journal of Chemical and Engineering Data*, *42*(4), 685–688. https://doi.org/10.1021/je970020p
9. Barden, D., & McGregor, L. (2017). A Guide to Modern Comprehensive Two-Dimensional Gas Chromatography. *The Column*, *13*(10), 14–20.

https://www.chromatographyonline.com/view/guide-modern-comprehensive-two-dimensional-gas-chromatography-0

10. Bouwer, E. J. (1997). Environmental organic chemistry. In *Journal of Contaminant Hydrology* (Vol. 25, Issues 1–2). https://doi.org/10.1016/s0169-7722(96)00030-7
11. Bowden, S. T., & Jones, W. J. (1948). XXII. Latent heat of vaporization and composition. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *39*(289), 155–161. https://doi.org/10.1080/14786444808521660
12. Chaurasia, G. (2017). *Effect of Acidic, Neutral, and Basic pH on Solubility and Partition-Coefficient of Benzoic Acid Between Water-Benzene System*. *8*(6), 2637–2640.
13. Chen, N. H. (1965). Generalized Correlation for Latent Heat of Vaporization. *Journal of Chemical and Engineering Data*, *10*(2), 207–210. https://doi.org/10.1021/je60025a047
14. Chiou, C. T., Freed, V. H., Schmedding, D. W., & Kohnert, R. L. (1977). Partition Coefficient and Bioaccumulation of Selected Organic Chemicals. *Environmental Science and Technology*, *11*(5), 475–478. https://doi.org/10.1021/es60128a001
15. Churchill, B., Acree, W. E., & Abraham, M. H. (2019). Development of Abraham model expressions for predicting the standard molar enthalpies of vaporization of organic compounds at 298.15 K. *Thermochimica Acta*, *681*. https://doi.org/10.1016/j.tca.2019.178372
16. Dexter, T. A. (2014). R: a language and environment for statistical computing. *Quaternary Research.*, *81*(2), 114–124. https://doi.org/10.1016/j.quascirev.2005.03.008
17. Fishtine, S. H. (1963a). Reliable latent heats of vaporization. *Industrial and Engineering Chemistry*, *55*(5), 49–54. https://doi.org/10.1021/ie50641a008
18. Fishtine, S. H. (1963b). RELIABLE LATENT HEATS OF VAPORIZATION. *Ind. Eng. Chem*, *55(4)*, 49–54.
19. Flôres, G. B., Staudt, P. B., & Soares, R. de P. (2016). Including dispersive interactions in the F–SAC model. *Fluid Phase Equilibria*, *426*, 56–64. https://doi.org/10.1016/j.fluid.2016.02.043
20. Franco, A., & Trapp, S. (2008). Estimation of the soil-water partition coefficient normalized organic carbon for ionizable organic chemicals. *Environmental Toxicology and Chemistry*, *27*(10), 1995–2004. https://doi.org/10.1897/07-583.1
21. Gharagheizi, F. (2012). Determination of normal boiling vaporization enthalpy using a

new molecular-based model. *Fluid Phase Equilibria*, *317*, 43–51. https://doi.org/10.1016/j.fluid.2011.12.024

22. Gharagheizi, F., Babaie, O., & Mazdeyasna, S. (2011). Prediction of vaporization enthalpy of pure compounds using a group contribution-based method. *Industrial and Engineering Chemistry Research*, *50*(10), 6503–6507. https://doi.org/10.1021/ie2001764
23. Giacalone, A. (n.d.). The invariance of heat of vaporization with temperature and the theory of corresponding states,. *Gazzetta Chimica Italiana*.
24. Goss, K. U., & Schwarzenbach, R. P. (1999). Empirical prediction of heats of vaporization and heats of adsorption of organic compounds. *Environmental Science and Technology*, *33*(19), 3390–3393. https://doi.org/10.1021/es980812j
25. Gouin, T., Armitage, J. M., Cousins, I. T., Muir, D. C. G., Ng, C. A., Reid, L., & Tao, S. (2013). Influence of global climate change on chemical fate and bioaccumulation: The role of multimedia models. *Environmental Toxicology and Chemistry*, *32*(1), 20–31. https://doi.org/10.1002/etc.2044
26. Holley, K., Acree, W. E., & Abraham, M. H. (2011). Determination of abraham model solute descriptors for 2-ethylanthraquinone based on measured solubility ratios. *Physics and Chemistry of Liquids*, *49*(3), 355–365. https://doi.org/10.1080/00319101003646553
27. Ivanciuc, O., Ivanciuc, T., Klein, D. J., Seitz, W. A., & Balaban, A. T. (2001). Wiener Index Extension by Counting Even/Odd Graph Distances. *Journal of Chemical Information and Computer Sciences*, *41*(3), 536–549. https://doi.org/10.1021/ci000086f
28. Kaminski, S., Kirgios, E., Bardow, A., & Leonhard, K. (2017). Improved Property Predictions by Combination of Predictive Models. *Industrial and Engineering Chemistry Research*, *56*(11), 3098–3106. https://doi.org/10.1021/acs.iecr.6b03125
29. Liu, Z. Y. (2001). Estimation of heat vaporization of pure liquid at its normal boiling temperature. *Chemical Engineering Communications*, *184*, 221–228. https://doi.org/10.1080/00986440108912849
30. Macleod, M., Scheringer, M., & Hungerbühler, K. (2007). Estimating enthalpy of vaporization from vapor pressure using Trouton’s rule. *Environmental Science and Technology*, *41*(8), 2827–2832. https://doi.org/10.1021/es0608186
31. Morgan, D. L., & Kobayashi, R. (1994). Extension of Pitzer CSP models for vapor pressures and heats of vaporization to long-chain hydrocarbons. *Fluid Phase Equilibria*,

94(C), 51–87. https://doi.org/10.1016/0378-3812(94)87051-9

32. Nabi, D. (2014). *Estimating Environmental Partitioning, Transport, and Uptake Properties for Nonpolar Chemicals Using GC× GC*. https://doi.org/10.5075/epfl-thesis-6013
33. Nabi, D., & Arey, J. S. (2017). Predicting Partitioning and Diffusion Properties of Nonpolar Chemicals in Biotic Media and Passive Sampler Phases by GC × GC. *Environmental Science and Technology*, *51*(5), 3001–3011. https://doi.org/10.1021/acs.est.6b05071
34. Nabi, D., Gros, J., Dimitriou-Christidis, P., & Arey, J. S. (2014). Mapping environmental partitioning properties of nonpolar complex mixtures by use of GC × GC. *Environmental Science and Technology*, *48*(12), 6814–6826. https://doi.org/10.1021/es501674p
35. Naef, R., & Acree, W. E. (2017). Calculation of five thermodynamic molecular descriptors by means of a general computer algorithm based on the group-additivity method: Standard enthalpies of vaporization, sublimation and solvation, and entropy of fusion of ordinary organic molecules and . *Molecules*, *22*(7). https://doi.org/10.3390/molecules22071059
36. Naseem, S., Zushi, Y., & Nabi, D. (2021). Development and evaluation of two-parameter linear free energy models for the prediction of human skin permeability coefficient of neutral organic chemicals. *Journal of Cheminformatics*, *13*(1), 1–23. https://doi.org/10.1186/s13321-021-00503-5
37. PITZER, K. S., LIPPMANN, D. Z., CURL, R. F., HUGGINS, C. M., & PETERSEN, D. E. (1993). *The Volumetric and Thermodynamic Properties of Fluids.: II. Compressibility Factor, Vapor Pressure and Entropy of Vaporization*. 303–310. https://doi.org/10.1142/9789812795960_0044
38. Poole, C. F., Ariyasena, T. C., & Lenca, N. (2013). Estimation of the environmental properties of compounds from chromatographic measurements and the solvation parameter model. *Journal of Chromatography A*, *1317*, 85–104. https://doi.org/10.1016/j.chroma.2013.05.045
39. Rebas, O., Boutra, B., Skander, N., & Chitour, C. E. (2016). Prédiction de l’enthalpie et de l’entropie de vaporisation normale par la méthode de contribution de groupes avec interactions des hydrocarbures purs, mélanges simples et fractions pétrolières. *Canadian Journal of Chemical Engineering*, *94*(1), 175–191. https://doi.org/10.1002/cjce.22343
40. Schenker, U., Macleod, M., Scheringer, M., & Hungerbühler, K. (2005). Improving data quality for environmental fate models: A least-squares adjustment procedure for

harmonizing physicochemical properties of organic compounds. *Environmental Science and Technology*, *39*(21), 8434–8441. https://doi.org/10.1021/es0502526

41. Schwarzenbach, R. P., Gschwend, P. M., & Imboden, D. M. (2002). *Environmental Organic Chemistry (2nd ed.)*.
42. Shanmugam, N., Eddula, S., Acree, W. E., & Abraham, M. H. (2021). Calculation of abraham model l-descriptor and standard molar enthalpies of vaporization for linear C7-C14 alkynes from gas chromatographic retention index data. *European Chemical Bulletin*, *10*(1), 46–57. https://doi.org/10.17628/ECB.2021.10.46-57
43. Sosnowska, A., Barycki, M., Jagiello, K., Haranczyk, M., Gajewicz, A., Kawai, T., Suzuki, N., & Puzyn, T. (2014). Predicting enthalpy of vaporization for persistent organic pollutants with quantitative structure-property relationship (QSPR) incorporating the influence of temperature on volatility. *Atmospheric Environment*, *87*, 10–18. https://doi.org/10.1016/j.atmosenv.2013.12.036
44. Tirumala, P., Huang, J., Eddula, S., Jiang, C., Xu, A., Liu, G., Acree, W. E., & Abraham, M. H. (2020). Calculation of abraham model l-descriptor and standard molar enthalpies of vaporization and sublimation for C9 - C26 mono-alkyl alkanes and polymethyl alkanes. *European Chemical Bulletin*, *9*(10–12), 317–328. https://doi.org/10.17628/ecb.2020.9.317-328

45. Vetere, A. (1979). New correlations for predicting vaporization enthalpies of pure compounds. *The Chemical Engineering Journal*, *17*(2), 157–162. https://doi.org/10.1016/0300-9467(79)85008-X
46. Vetere, A. (1995). Methods to predict the vaporization enthalpies at the normal boiling temperature of pure compounds revisited. *Fluid Phase Equilibria*, *106*(1–2), 1–10. https://doi.org/10.1016/0378-3812(94)02627-D
47. Wadsö, I., Murto, M.-L., Bergson, G., Ehrenberg, L., Brunvoll, J., Bunnenberg, E., Djerassi, C., & Records, R. (1966). Heats of Vaporization for a Number of Organic Compounds at 25 degrees C. *Acta Chemica Scandinavica*, *20*, 544–552. https://doi.org/10.3891/acta.chem.scand.20-0544
48. Wagner, W. (1973). New vapour pressure measurements for argon and nitrogen and a

new method for establishing rational vapour pressure equations. *Cryogenics*, *13*(8), 470–482. https://doi.org/10.1016/0011-2275(73)90003-9

49. Walton, J. (1989). Vapour pressures up to their critical temperatures of normal alkanes and 1-alkanols. *Pure and Applied Chemistry*, *61*(8), 1395–1403. https://doi.org/10.1351/pac198961081395
50. Wang, R. Y., & Shi, J. (1990). A reference value method of the corresponding states principle for the prediction of latent heats of vaporization of pure liquids. *Thermochimica Acta*, *169*(C), 239–246. https://doi.org/10.1016/0040-6031(90)80151-N
51. Wright, F. J. (1960). Latent heat of vaporisation and composition. *Recueil Des Travaux Chimiques Des Pays-Bas*, *79*(8), 784–789. https://doi.org/10.1002/recl.19600790803
52. Wu, J. (2020). Soil-Air Partition Coefficients of Persistent Organic Pollutants Decline from Climate Warming: a Case Study in Yantai County, Shandong Province, China. *Water, Air, and Soil Pollution*, *231*(7). https://doi.org/10.1007/s11270-020-04718-4
53. Zhao, L., Li, P., & Yalkowsky, S. H. (1999). Predicting the Entropy of Boiling for Organic Compounds. *Journal of Chemical Information and Computer Sciences*, *39*(6), 1112–1116. https://doi.org/10.1021/ci990054w
54. Zhao, L., Ni, N., & Yalkowsky, S. H. (1999). A modification of Trouton's rule by simple molecular parameters for hydrocarbon compounds. *Industrial and Engineering Chemistry Research*, *38*(1), 324–327. https://doi.org/10.1021/ie9803570
55. Goss, K. U. (2005). Predicting the equilibrium partitioning of organic compounds using just one linear solvation energy relationship (LSER). *Fluid Phase Equilibria*, *233*(1), 19–22. https://doi.org/10.1016/j.fluid.2005.04.006
56. Goss, K. U., & Schwarzenbach, R. P. (2001). Linear free energy relationships used to evaluate equilibrium partitioning of organic compounds. In *Environmental Science and Technology* (Vol. 35, Issue 1, pp. 1–9). https://doi.org/10.1021/es000996d
57. Nedyalkova, M., Madurga, S., Tobiszewski, M., & Simeonov, V. (2019). Calculating the Partition Coefficients of Organic Solvents in Octanol/Water and Octanol/Air. *Journal of chemical information and modeling*
58. Mintz, C., Clark, M., Acree, W. E., & Abraham, M. H. (2006). Enthalpy of Solvation Correlations for Gaseous Solutes Dissolved in Water and in 1-Octanol Based on the Abraham Model. *Journal of Chemical Information and Modeling, 47*(1), 115-121.

doi:10.1021/ci600402n

59. Goss, K.-U., & Schwarzenbach, R. P. (2003). Rules of Thumb for Assessing Equilibrium Partitioning of Organic Compounds: Successes and Pitfalls. *Journal of Chemical Education, 80*(4), 450. doi:10.1021/ed080p450
60. Gawor, A., & Wania, F. (2013). Using quantitative structural property relationships, chemical fate models, and the chemical partitioning space to investigate the potential for long 38 range transport and bioaccumulation of complex halogenated chemical mixtures.
61. *Environmental Science: Processes & Impacts, 15*(9), 1671-1684. doi:10.1039/c3em00098b
62. US EPA. (2016). *Estimation Programs Interface SuiteTM for Microsoft Windows, v 4.00 (KowWIN, ver. 1.68).*
63. Van Noort, P. C. M. (2013). A possible simplification of the Goss-modified Abraham solvation equation. *Chemosphere*, *93*(9), 1742–1746. https://doi.org/10.1016/j.chemosphere.2013.05.081
64. XLSTAT, A. (2020). *Data Analysis and Statistics Software for Microsoft Excel; Addinsoft:Paris,.*
65. Zhao, Y. H., & Abraham, M. H. (2005). Octanol/water partition of ionic species, including 544 cations. *Journal of Organic Chemistry, 70*(7), 2633–2640. https://doi.org/10.1021/jo048078b
66. Core, R., & Team, R. (2020). *A language and environment for statistical computing, R Foundation for Statistical Computing; 2013*. https://www.r-project.org/.
67. Altomare, C., Cellamare, S., Carotti, A., & Ferappi, M. (1994). Linear solvation energy relationships in reversed-phase liquid chromatography. Examination of RP-8 stationary phases for measuring lipophilicity parameters. *Farmaco*, *49*(6), 393–401.
68. Abraham, Michael H. (1993). Scales of solute hydrogen-bonding: Their construction and application to physicochemical and biochemical processes. *Chemical Society Reviews, 22*(2), 73–83. https://doi.org/10.1039/CS9932200073
69. Abraham, Michael H., & Acree, W. E. (2009). Prediction of convulsant activity of gases and vapors. *European Journal of Medicinal Chemistry*, *44*(2), 885–890. https://doi.org/10.1016/j.ejmech.2008.05.027
70. Chickos, J. S., Hosseini, S., & Hesse, D. G. (1995). Determination of vaporization enthalpies of simple organic molecules by correlations of changes in gas

chromatographic net retention times. *Thermochimica Acta*, *249*(C), 41–62. https://doi.org/10.1016/0040-6031(95)90670-3

71. Ellison, H. R. (2005). Enthalpy of Vaporization by Gas Chromatography. A Physical Chemistry Experiment. *Journal of Chemical Education*, *82*(7), 1086. https://doi.org/10.1021/ed082p1086
72. Gobble, C., & Chickos, J. S. (2015). A Comparison of Results by Correlation Gas Chromatography with Another Gas Chromatographic Retention Time Technique. the Effects of Retention Time Coincidence on Vaporization Enthalpy and Vapor Pressure. *Journal of Chemical and Engineering Data*, *60*(9), 2739–2748. https://doi.org/10.1021/acs.jced.5b00444.

Supporting Information

Supporting information of my thesis comprises of all tables, plots, figures and other relevant detailed description of analytical tools.
Supporting information of my thesis can be access by the following link:

https://docs.google.com/document/d/1IWHE6ZqdJkAQg_6AbaOlmPwsI8ys-TL8/edit?usp=sharing&ouid=112805291625940808376&rtpof=true&sd=true