Available online at www.sciencedirect.com
ScienceDirect
Journal of Hazardous Materials 147 (2007) 424–430
Journal of Hazardous Materials
www.elsevier.com/locate/jhazmat

# Quantitative structure–property relationship studies for predicting flash points of alkanes using group bond contribution method with back-propagation neural network

Yong Pan *, Juncheng Jiang, Zhirong Wang

*Institute of Safety Engineering, Nanjing University of Technology, Nanjing 210009, China*

Received 20 November 2006; received in revised form 6 January 2007; accepted 8 January 2007
Available online 12 January 2007

**Abstract**

Models of relationships between structure and flash point of 92 alkanes were constructed by means of artificial neural network (ANN) using group bond contribution method. Group bonds were used as molecular structure descriptors which contained information of both group property and group connectivity in molecules, and the back-propagation (BP) neural network was employed for fitting the possible nonlinear relationship existed between the structure and property. The dataset of 92 alkanes was randomly divided into a training set (62), a validation set (15) and a testing set (15). The optimal condition of the neural network was obtained by adjusting various parameters by trial-and-error. Simulated with the final optimum BP neural network [9-5-1], the results showed that the predicted flash points were in good agreement with the experimental data, with the average absolute deviation being 4.8 K, and the root mean square error (RMS) being 6.86, which were shown to be more accurate than those of the multilinear regression method. The model proposed can be used not only to reveal the quantitative relation between flash points and molecular structures of alkanes, but also to predict the flash points of alkanes for chemical engineering.


*Keywords:* Quantitative structure–property relationship (QSPR); Back-propagation (BP) neural network; Flash point; Group bond contribution method; Alkane

## 1. Introduction

The flash point of a flammable (or combustible) compound is the temperature at which the vapor pressure of the substance is such as to produce a concentration of vapor in the air that corresponds to the lower flammable limit [1]. This parameter provides the knowledge necessary for understanding the fundamental physical and chemical processes of combustion. Moreover, it is of importance in practice for safety considerations in storage, processing, and handling of a given compound, and is one of the major flammability characteristics used to assess the fire and explosion hazards of chemical compounds.

Reliable values of flash points are always desirable, and some of them can be measured by two currently-accepted experimental methods, which are the closed cup test and the open cup test [2]. However, for many other compounds, the experimental flash point values are scarce or too expensive to obtain. What is more, for toxic, volatile, explosive, and radioactive compounds, the experimental determination of flash point values is more difficult. Hence the development of estimation methods which are desirably convenient for predicting the flash points in short is required.

Quantitative structure–property relationships (QSPR) method which relates descriptors of the molecular structure to the properties of chemical compounds, has been reported quite extensively in the literature for the prediction of flash point [3–10]. For example, the first method for estimating the flash point of organic compounds from their molecular structure was developed by Suzuki et al. The 25 atomic and group contributions were employed for predicting the flash points of 33 aliphatic and 26 aromatic hydrocarbons with an average absolute deviation of 12.2 and 6.1 °C, respectively. The average deviation for the 59 hydrocarbon compounds tested was 9.5 °C. In another work, Tetteh et al. used a radial basis function neural network for the estimation of flash points for a large set of 400 compounds from different classes. The structures were described simply with a molecular connectivity index and counts of the 25 functional groups present in the molecules.

* Corresponding author. Tel.: +86 25 83587305; fax: +86 25 83587411.
*E-mail address:* yongpannjut@163.com (Y. Pan).

0304-3894/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhazmat.2007.01.025

The average absolute error for the test set in flash point prediction was 11.9 °C using a RBFNN with a 26-36-2 configuration. Katritzky et al. studied quantitative structure–flash point relationships for a diverse set of 271 compounds. The general three-parameter QSPR model provided $R^2 = 0.9020$ and $s = 16.1$ K. When the boiling point was used as a descriptor in the model, the correlation was improved to $R^2 = 0.9529$. Meanwhile, the study on mixture flash points [9,10] which can display non-ideal behavior with important safety consequences has also been developed. Liaw et al. proposed a mathematical model, which could be used for predicting the flash point of aqueous-organic solutions, and the results revealed that this model was able to precisely predict the flash point over the entire composition range of binary aqueous-organic solutions by way of utilizing the flash point data pertaining to the flammable component.

The group bond contribution method was recently proposed by Wang [11] for the description of molecules structure. This method combined together information of both group property and connectivity in the analyzed molecules, and has been successfully used in the estimation of physical and chemical properties, such as density [11] and boiling point [12].

In recent years, the modeling technique of artificial neural network (ANN) has been widely used in the field of QSPR [4,8,13,14]. ANN is a powerful tool for correlating and estimating chemical properties and one of a group of intelligence technologies for data analysis that differ from other classical analysis techniques. The advantage of ANN is in its inherent ability to incorporate nonlinearity and cross-product terms into the model. Besides, it is also able to acquire an estimate function from studied samples while the form of the mathematical function is unknown.

In this paper, we developed a method to estimate the flash points of 92 alkanes based on the back-propagation (BP)-ANN using group bond contribution method. Group bonds which obtained from chemical structure are used as molecular structure descriptors, and these descriptors are quantitatively related to flash points of 92 alkanes by BP neural network.

## 2. Group bond contribution method

One of the most widely used methods proposed for prediction of properties from molecular structure is group contribution method. It is based on the assumption that the contribution of a certain group is completely the same in different molecules, and the properties of compounds are considered as the contribution addition of groups which constituted the compounds. The group contribution method works well for a large number of compounds, however, difficulties may arise in decomposing some structures into appropriate groups whose constants are available. Several correction factors are also needed for some molecular interactions, for group contribution method takes into account only the contribution of groups in the molecule but the interaction between groups and chemical bonds. Besides, group contribution method has a weak ability for distinguishing the isomeric compounds. For instance, the structure difference of 2-methylhexane, 3-methyhexane and 3-ethylpentane cannot be distinguished from only molecular groups, because there are 3 "–$CH_3$"group, 3 "–$CH_2$–" group and 1 "≺CH–" group in each of the three compounds above, while the flash point values of them were 269, 258 and 255 K, respectively [15]. Because of such causations, group contribution method has some limitations in QSPR studies for property calculation.

Table 1
Group bonds presented in the alkane molecules

| No. | Group bond | No. | Group bond | No. | Group bond |
|---|---|---|---|---|---|
| 1 | $CH_3$–$CH_2$– | 4 | –$CH_2$–$CH_2$– | 7 | ≻CH–CH≺ |
| 2 | $CH_3$–CH≺ | 5 | –$CH_2$–CH≺ | 8 | ≻CH—C— |
| 3 | $CH_3$—C— | 6 | —$CH_2$—C— | 9 | —C—C— |

However, the group bond contribution method can cover these shortages above. Group bond which is defined as an integration of two molecular groups and the chemical bond between them, contains information of both group and chemical bond. Strict quantitative relationships among the number of group bond, number of group and number of chemical bond are existed, and the number of group and chemical bond can be confirmed from the number of group bond in molecule, but the species and number of group bond cannot be ascertained from group and chemical bond. Thus the group bond contribution method contains both group contribution and bond contribution, and could have better and more comprehensive prediction abilities than the group contribution method to a certain extent. Besides, the group bond contribution method takes into account both group property and connectivity in the analyzed molecules, so it may have a great ability for the identification of isomeric compounds.

In this work, the flash point values of alkanes were regarded as the concerted contribution of numerous group bonds constituting the alkane molecules. There were four species of groups (–$CH_3$, ≻CH–, –$CH_2$–, ≻C≺) existed in the molecular structure of alkanes, which constituted a set of nine species of group bonds presented in all the alkane molecules except ethane. All the nine group bonds were listed in Table 1.

Furthermore, because of the possibility that highly nonlinear interaction may exist among the group bonds in the chemical molecules, the contribution value of each group bond obtained by mathematical regression method could not satisfactorily show the difference of contributions to property for one certain group bond in different molecules. However, ANN could describe such nonlinear interaction between the group bonds satisfactorily for its inherent ability of nonlinear fitting. So in this work we combined together both group bond contribution method and ANN for the flash point prediction of alkanes.

## 3. Experiment

### 3.1. Data sets

The applicability and accuracy of a flash point estimation model are directly affected by the size and quality of the training