胡之德

科学论文选集（下）

《胡之德科学论文选集》编辑组　编

兰州大学出版社
LANZHOU UNIVERSITY PRESS

胡之德

科学论文选集（下）

《胡之德科学论文选集》编辑组　编

兰州大学出版社

图书在版编目(CIP)数据

胡之德科学论文选集/《胡之德科学论文选集》编辑组编. —兰州:兰州大学出版社,2006.9

ISBN 7-311-02891-4

Ⅰ.胡… Ⅱ.胡… Ⅲ.自然科学—文集

Ⅳ.N53

中国版本图书馆 CIP 数据核字(2006)第 117739 号

胡之德科学论文选集

《胡之德科学论文选集》编辑组 编

兰州大学出版社出版发行

兰州市天水南路 222 号 电话:8912613 邮编:730000

E-mail:press@onbook.com.cn

http://www.onbook.com.cn

兰州大学出版社激光照排中心排版

兰州人民印刷厂印刷

开本:787×1092 1/16 印张:100.75

2006 年 9 月第 1 版 2006 年 9 月第 1 次印刷

插页:5 字数:2317 千字

ISBN7-311-02891-4/O·197 定价:150.00 元
(上、下册)

目录

TABLE OF CONTENTS

第一部分 分离科学

第二部分 化学信息学

第三部分 生命分析化学

第四部分 荧光分析及其它

QSAR Comb Sci 2003,22: 29-48

Quantitative structure property relationship models for the prediction of liquid heat capacity

Yao Xiaojun [1,2], Fan Botao [1]*, Doucet J. P.,[1] A. Panaye[1], Liu Mancang [2], Zhang Ruisheng [2], Zhang Xiaoyun [2], Hu Zhide [2]*

[1]Université Paris 7–Denis Diderot, ITODYS–CNRS UMR 7086, 1, Rue Guy de la Brosse, 75005 Paris, France

[2]Department of Chemistry, Lanzhou University, Lanzhou 730000, China

Abstract: Quantitative Structure-Property Relationship (QSPR) models based on molecular descriptors derived from molecular structures have been developed for the prediction of liquid heat capacity at 25 °C using a diverse set of 871 organic compounds. The molecular descriptors used to represent molecular structures include constitutional and topological indices and quantum chemical parameters. Forward stepwise regression and radial basis function neural networks (RBFNNs) were used to construct the QSPR models. The root mean square errors in liquid heat capacity predictions for the training, test and overall data sets are 16.857, 18.744 and 17.141 heat capacity units, respectively. The prediction results are in agreement with the experimental values, but the RBFNN model seems to be better than stepwise regression method.

1. Introduction

The study of the quantitative relationship between property/activity and molecular structure (QSPR/QSAR) is an important research area in computational chemistry and has been widely used in the prediction of physicochemical properties and biological activities of organic compounds[1-2]. This kind of study can not only develop a method for the prediction of the property under investigation of new compounds that have not been synthesized but also can identify and describe important structural features of molecules that are relevant to variations in molecular properties, thus gain some insight into structural factors affecting molecular properties. To develop a QSPR model, the following steps are usually involved: data collection, molecular geometry optimization, molecular descriptors generation, descriptors selection, model development and finally model performance evaluation.

One of the important problems in QSPR is the description of molecular structures using molecular descriptors, which can include structural information as much as possible. At present, there exists a great number of molecular descriptors that encode constitutional, topological, geometry and electronic features of organic compounds[3-5]. Among various structure descriptors,

those derived from molecular structure alone have a particular advantage of the possibility to calculate them based only on molecular structural feature and to be applicable to different families of compounds[6-7]. After the calculation of molecular descriptors, linear methods, such as multiple linear regression (MLR), principal component regression (PCR) and partial least squares (PLS) or non-linear methods, such as many types of artificial neural networks, can be used in the development of a mathematical relationship between the structural descriptors and the property to be predicted.

Physical and thermodynamic property data of organic compounds such as liquid heat capacity are important in the engineering design and operation of industrial chemical processes[8]. In liquid phase reactions, the liquid heat capacity is often required to evaluate the energy variation for the liquid chemical reactants involved in the reaction. The heat capacity of a compound is defined as the amount of energy needed to raise a unit of the substance in question by one unit of temperature. The liquid heat capacity data is also needed in the heat exchanger and energy balance design calculations. Since the experimental determination of heat capacity is both time-consuming and expensive, and there is increased need of reliable physical and thermodynamic data for the optimization of chemical processes, it would be very useful to develop predictive models that can be used to predict these properties of organic compounds that are not synthesized or their properties are unknown.

Although many reported QSPR methods have been successfully used to predict a diverse set of physicochemical properties, their use in predicting heat capacity is rather limited[9-10]. Gakh et al obtained a QSPR model for heat capacity of alkanes using graph theory descriptors and neural networks with a RMS of 4.04 heat capacity unit[9]. Liu et al developed several QSPR models of alkanes using their molecular electronic edge vectors (MEDV). The RMS for the heat capacity of 134 alkanes was 3.81 heat capacity unit[10]. The works reported are only limited to the investigations of alkanes. In our previous works we have successfully developed several QSPR models based on RBFNNs for the prediction of physicochemical properties[11-15] and HPLC retention indices of N-Benzylideneanilines[16]. The goal of the present study is to extend our previous investigations in order to, for the first time, establish a QSPR model that can predict the liquid heat capacity at 25°C for a diverse set of organic compounds dependent only upon their molecular structures. MLR and RBFNNs are applied to establish quantitative linear and non-linear relationships between heat capacity and molecular descriptors respectively. The data set used in our work is more diverse and the models developed are more general and practical with respect to the models reported early by other authors.

2. Materials and methods

2.1 Data set

The data used in this investigation were collected from literature[8]. The set consists of 871

organic compounds. The data set includes hydrocarbons, fluorocarbons, chlorocarbons, bromocarbons, iodocarbons, alcohols, acids, ketones, aldehydes, ethers, esters, amines, nitriles, etc. A complete list of the compound names and their corresponding experimental liquid heat capacities at 25°C is given in Table 1. The data were randomly separated into two subsets: a training set of 746 compounds and a test set of 125 compounds. The training set was used to adjust the parameters of RBFNNs and the test set was used to evaluate the predictive ability of RBFNNs.

Table 1 The compounds and the predicted results of liquid heat capacities: (J/mol.K)

Number	NAME	HEATCAP	MLR	RBFNNs
1	Bromochlorodifluoromethane	127.81	106.55	127.23
2*	Bromotrichloromethane	130.65	108.16	130.03
3	Bromotrifluoromethane	143.25	104.50	126.00
4	dibromodifluoromethane	129.94	109.53	129.03
5	dichlorodifluoromethane	121.07	105.12	125.43
6	phosgene	112.03	118.18	117.02
7	trichlorofluoromethane	68.65	108.15	128.95
8	carbontetrachloride	130.72	100.65	126.01
9*	tribromomethane	129.94	141.93	136.20
10	chlorodifluoromethane	111.66	100.40	105.96
11	dichlorofluoromethane	107.20	104.37	106.08
12	chloroform	112.49	97.81	99.01
13	bromochloromethane	100.75	113.78	100.37
14	dibromomethane	105.11	129.51	116.57
15	dichloromethane	101.98	95.29	82.05
16*	difluoromethane	107.05	101.92	94.33
17	diiodomethane	135.45	145.06	141.58
18	formic acid	98.40	98.60	106.48
19	methylbromide	85.16	119.67	81.05
20	methylchloride	82.33	104.69	61.86
21	methyliodide	82.91	128.28	98.97
22	nitromethane	104.22	111.55	127.14
23*	methanol	79.93	123.08	94.04
24	methylmercaptan	96.39	131.43	92.57
25	methylamine	114.19	120.29	99.43
26	bromotrifluoroethylene	135.33	135.29	140.92
27	1,2-dibromotetrafluoroethane	184.03	153.13	179.99
28	chlorotrifluoroethylene	150.36	135.29	140.92
29	chloropentafluoroethane	174.61	150.71	173.07

Table 1 (continued)

Number	NAME	HEATCAP	MLR	RBFNNs
30*	1,2-dichlorotetrafluoroethan	110.13	150.31	174.06
31	1,1,2-trichlorotrifluoroetha	169.44	153.97	177.66
32	tetrachloroethylene	146.93	134.64	140.24
33	trichloroacetylchloride	168.81	161.40	169.97
34	2-chloro-1,1-difluoroethylen	126.95	140.04	138.28
35	trichloroethylene	123.67	130.81	128.04
36	dichloroacetyl chloride	154.23	152.81	152.72
37*	trichloroacetaldehyde	153.22	150.51	154.35
38	pentachloroethane	190.91	146.28	163.41
39	trifluoroacetic acid	170.15	155.99	160.88
40	pentafluoroethane	168.85	143.59	157.87
41	1,1,2,2-tetrabromoethane	168.43	157.56	156.62
42	1,1-dichloroethylene	112.43	122.69	121.29
43	cis-1,2-dichloroethylene	117.88	116.74	110.47
44*	trans-1,2-dichloroethylene	118.37	115.54	109.88
45	chloroacetyl chloride	142.92	138.03	137.11
46	dichloroacetaldehyde	134.07	141.23	138.66
47	dichloroacetic acid	172.72	164.08	161.64
48	1,1,1-trichlorofluoroethane	151.42	147.95	154.60
49	1,1,1,2-tetrachloroethane	154.63	144.61	151.58
50	1,1,2,2-tetrachloroethane	159.39	133.99	139.18
51*	1,1,1,2-tetrafluoroethane	141.80	135.99	145.90
52	vinyl bromide	103.13	109.82	106.96
53	vinyl chloride	86.01	106.30	101.32
54	1-chloro-1,1-difluoroethane	132.37	127.27	140.17
55	acetyl chloride	115.02	114.12	121.16
56	chloroacetaldehyde	115.19	124.99	124.32
57	methyl chloroformate	141.03	114.12	121.16
58*	1,1,1-trichloroethane	145.56	137.87	147.19
59	1,1,2-trichloroethane	151.46	131.14	130.57
60	1,1,1-trifluoroethane	139.86	122.51	137.18
61	acetonitrile	91.94	99.29	88.93
62	1,1-dibromoethane	133.61	127.47	128.10
63	1,2-dibromoethane	131.75	125.05	122.86
64	1,1-dichloroethane	131.00	128.86	128.18