Microporous and Mesoporous Materials 132 (2010) 246–257

Contents lists available at ScienceDirect

Microporous and Mesoporous Materials

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

Statistical analysis and Partial Least Square regression as new tools for modelling and understanding the adsorption properties of zeolites

P. Leflaive [a,*], G.D. Pirngruber [a], A. Faraj [b], P. Martin [c], G.V. Baron [c], J.F.M. Denayer [c]

[a] Separation Department, IFP-Lyon, BP3, 69360 Solaize, France
[b] Applied Mathematics Department, IFP, 1&4, Avenue de Bois-Préau, 92852 Rueil-Malmaison Cedex, France
[c] Department of Chemical Engineering, Vrije Universiteit Brussel, Pleinlaan 2, B-1050 Brussels, Belgium

ARTICLE INFO

Article history:
Received 6 October 2009
Received in revised form 25 February 2010
Accepted 2 March 2010
Available online 4 March 2010

Keywords:
Beta
Mordenite
ZSM-5
ZSM-22
Faujasite

ABSTRACT

Statistical data analysis was applied to examine previously published adsorption enthalpies of linear and branched alkanes on zeolites Na-Y, Na-USY, Mordenite, Beta, ZSM-5 and ZSM-22. The objective was to establish predictive models that correlate the adsorption enthalpies for the six zeolites to simple structural properties or descriptors of the adsorbates. Two sets of descriptors were tested; a first set was chosen from a library of topological descriptors (the Wiener index, the Kier shape indices and the Kier + Hall connectivity indices), the other set was custom-made, by directly parametrizing the number of carbon atoms as well as position and degree of branching. From the set of 17 available adsorbates, training sets of 10 molecules were selected by a clustering analysis and this training set was subjected to a Partial Least Squares (PLS) regression in order to establish a linear relationship between the adsorption enthalpy and the set of descriptors. The model was then applied to predict the adsorption enthalpies of the other molecules. In all cases, predicted and experimental values were in good agreement. The model based on the custom-made descriptors that code the branching of the molecules was less precise, but allowed a more straightforward interpretation of the coefficients of the model, which provides very interesting insights into the adsorption behaviour of the linear and branched alkanes on large and medium pore zeolites.

1. Introduction

Zeolites are widely used as catalyst or separating agents in the refining and petrochemical industries. In the field of catalysis, the understanding of adsorption phenomena is critical in the comprehension and the modelling of reaction kinetics on zeolites. It can indeed play a key role in product distribution and catalyst selectivity [1]. Selectivity and adsorption efficiency are often characterized by the Henry adsorption constants and the adsorption enthalpies at low-coverage. The catalytic or separating performance is then correlated to the adsorbate–adsorbent interaction via one of these or both parameters. Characterizing the adsorption enthalpies is, therefore, a crucial step in catalyst and adsorbent development.

The number of zeolite topologies referenced by the International Zeolite Association is increasing with the rapid discovery of new structures. Testing a large number of different zeolites for a given reaction or separation requires a huge experimental effort. It would therefore be advantageous, for a preliminary evaluation, if one would be able to estimate in advance the adsorption properties of zeolites for various hydrocarbons or at least to quickly evaluate the potential of a zeolite as catalyst or adsorbent. Nevertheless, even if different attempts were made to correlate adsorption properties and zeolite structures using molecular modelling (recently reviewed by Smit [2]) or other techniques (see for example [3,4]), no model can efficiently be used to predict the adsorption properties of a wide variety of zeolites and rapidly forecast the properties of a new structure. The experimental measurement of Henry constants, low-coverage adsorption enthalpies or complete adsorption isotherms of the hydrocarbons on different zeolites is therefore still necessary.

High throughput equipments as, for example, parallel reactors can now carry out many experiments per day such as reactions [5,6] or breakthrough tests [7], and yield a large amounts of results. Some of these high throughput equipments are specifically designed for zeolite catalysts and adsorbents preparation and screening [7–10]. Nevertheless, the number of potential catalysts/adsorbents is much too large for exhaustive screening, even using High Throughput Screening (HTS) techniques. Thus, the use of predictive modelling, using methods such as quantitative structure–activity relationship (QSAR) is required to reduce the catalyst or adsorbent space through statistical data analysis and experimental design. A model that fits the experimental data successfully can, in some cases, also provide a better understanding of the adsorption

* Corresponding author. Tel.: +33 4 78 02 28 34; fax: +33 4 78 02 20 66.
E-mail address: philibert.leflaive@ifp.fr (P. Leflaive).

1387-1811/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.micromeso.2010.03.004

phenomenon. For this reason, predictive methods for estimating key properties such as adsorption enthalpies from (high throughput) experimental data combined with easily available properties (descriptors) of the adsorbates are highly desirable.

Several examples of such predictive models already exist in the literature. For example, Wu et al. proposed a predictive model for the affinity coefficient of the Dubinin–Radushkevich equation which expresses the ratio of the characteristic free energies of adsorption of a gaseous compound and a reference vapor on an activated carbon [11]. This parameter is independent of the adsorbent (activated carbon) and a function of properties of the adsorbate only. The training set (eight compounds) of the model was first selected from a set of 68 organic chemicals with the help of PCA and statistical design. Partial Least Square (PLS) analysis was then used to correlate the measured affinity coefficients of the training set compounds with their physico-chemical properties (45 parameters). The final model was based only on three parameters, the molecular weight, the van der Waals volume of the compound and the energy of interaction between the compound and a graphite model surface, which according to the authors demonstrates the importance of adding a variable describing the energy of adsorption to the carbon surface to the size and shape descriptors.

Early works of Nirmalakhandan et al. [12,13] used a QSAR technique based on Polanyi–Radushkevich correlations to study the adsorption of organic vapors on activated carbon. In these studies, two QSAR approaches were used: one based on the molar volume, the other using the valence modified first-order molecular connectivity index $^{1}\chi^{v}$ which is a topological molecular descriptor calculated from the skeleton of the molecule as described in [13]. The latter approach has the advantage that no external input of experimental data is required as $^{1}\chi^{v}$ can easily be calculated from the molecular structure, without recourse to often incomplete descriptor databases, additional experimental measurements or computations. It moreover yields better results than the first approach, indicating that efficient prediction of adsorption properties can be obtained using topological molecular descriptors.

Brasquet et al. also investigated the possibility to set up a quantitative relationship between the strength of adsorption of a wide range of organic compounds in water and their molecular structure represented by some of their molecular connectivity indices (MCI) [14,15]. Brasquet and Le Cloirec studied the adsorption properties of 368 aliphatic and aromatic compounds on an activated carbon sample and of 55 compounds on activated carbon fibers [15]. In both cases, experimental data were fitted with the Freundlich equation. The obtained Freundlich parameter K was used in the correlation models. Multiple linear regression (MLR) models were found to be efficient to represent the training data and to predict the adsorption data of compounds out of the training database.

More recently the same group used a multi linear regression model to study the integral adsorption enthalpies from zero coverage to saturation of volatile organic compounds (VOCs) on activated carbons [16]. The MLR model was used to discriminate the significant input variables from a set of eight parameters for the molecules and five for the activated carbons. Explicative variables were found to be the polarizability, the heat of vaporization, the ionization potential and the surface tension for the VOCs while only the average micropore radius was retained for the activated carbons.

A particular approach was developed by Meeks and Rybolt to describe the adsorption of a large range of molecules on representative adsorbents (zeolite, silica gel, carbon black, microporous carbon) [17]. Each gas molecule is considered as a series of structural units ($-CH_3$, $-CH_2$, $-CH$ groups for example) that make a fixed contribution to the total molecular adsorption energy. In some cases, a coefficient related to the dipole moment of the molecule is added. Multiple regression analyses yielded a correlation coefficient of 0.99 for all considered gas–solid systems. This structural method was shown to be superior to correlations based on descriptors such as boiling points, critical constants, partition coefficients, and various combinations of other physical properties that were also examined. However, this method has limitations as it cannot discriminate isomers such as 2- and 3-methylpentane or 2-, 3- and 4-methylheptane, which typically have a different enthalpy of adsorption in microporous adsorbents.

A so-called "additivity method" was used by Narasimham et al. [3] to estimate the pore mouth physisorption of alkanes on ZSM-22. Due to the very narrow channel structure (0.44×0.55 nm) of this 10-membered ring zeolite, *n*-alkanes can adsorb into the micropores while the iso-alkanes adsorb at the pore mouths only, resulting in a very specific adsorption behaviour. For the iso-alkanes, multiple modes of physisorption exist at the pore mouths, in which the different "straight ends" of the iso-alkanes protrude into the micropores. The standard physisorption enthalpy for each of the physisorption modes was computed using an additivity principle. The six parameters of the model were estimated from the regression of the total set of experimental data, i.e. the physisorption properties of normal and branched alkanes with 5–9 carbon numbers on ZSM-22. Composite standard physisorption enthalpies were then calculated and compared to the experimental data. A good agreement was found between calculated and experimental values except for a few multi-branched molecules such as 2,2-dimethylbutane, 2,3-dimethylpentane and 2,5-dimethylhexane for which the calculated values are found to underpredict the experimental data of more than 10 kJ mol^{-1}. The model could nevertheless well describe the variations of the iso-alkanes as compared to the corresponding *n*-alkane and moreover seems to give better results than previously reported configurational bias Monte Carlo calculations [18].

In summary, works published in the literature related to quantitative structure properties relationships (QSPR) methods used to predict adsorption properties of molecules on different solids reveal that methods or models based on structural information yield better results than those correlating the adsorption properties to a large array of other molecular descriptors. Predictive methods for estimating adsorption enthalpies should therefore preferably be based on molecular structure [19].

The purpose of this work is to use a descriptor based QSAR approach to analyse and predict the adsorption properties of hydrocarbons on zeolites with different topology. In particular, the models are used to predict the zero coverage adsorption enthalpies of 17 alkanes and iso-alkanes on zeolites Na-Y, Na-USY, Beta, Mordenite, ZSM-5 and ZSM-22. The experimental data were published before [20]. Two different sets of descriptors, both based on topological and structural information about the adsorbate molecules, are used. Two PLS models, one based on each descriptors set, were applied to predict the adsorption enthalpy at zero coverage.

2. Methods

2.1. Principal Components Analysis of the adsorption data

Before introducing descriptors and modelling of the experimental data, a multidimensional exploratory data analysis of the response variables (i.e. the experimental adsorption enthalpy data in Table 1) was carried out using Principal Component Analysis (PCA) [21].

Principal Component Analysis (PCA) is a mathematical method for multidimensional data analysis which consists in seeking the directions of space which represent best the correlations between