Available online at www.sciencedirect.com
ScienceDirect
Chemical Physics Letters 433 (2006) 125–129
CHEMICAL PHYSICS LETTERS
www.elsevier.com/locate/cplett

A new approach to the characterization of nanomaterials: Predicting Young's modulus by correlation weighting of nanomaterials codes

Andrey A. Toropov *, Jerzy Leszczynski

Computational Center for Molecular Structure and Interactions, Jackson State University, Jackson, MS, USA

Received 27 September 2006; in final form 30 October 2006
Available online 10 November 2006

Abstract

Twenty-nine nanomaterials (atom composition and conditions of syntheses) have been studied. Their characteristics were applied to predictions of Young's modulus values (in GPa). The obtained statistical characteristics of the models are reasonably good, $n = 21$, $r^2 = 0.9757$, $s = 18.3$ GPa, $F = 761$ (training set) and $n = 8$, $r^2 = 0.8952$, $r^2_{\text{pred}} = 0.8880$, $s = 34.7$ GPa, $F = 51$ (test set).

1. Introduction

Nanomaterials are becoming an important component of modern life and are the subject of many investigations in various areas of natural sciences. However, theoretical modeling of physicochemical and biological activity of these species is very scarce. It is a well-known to predict the properties and/or activities of 'classical' substances via correlating with some molecular descriptors. These methods are often cited in the literature as quantitative structure–property/activity relationships (QSPR/QSAR). The prediction of properties/activities by QSPR/QSAR is based on information concerning the molecular structure of the molecules of interest. As a rule the molecular graph is an elucidation of molecular structure in the QSPR/QSAR analysis [1–4]. As an alternative to the molecular graph in QSPR/QSAR analyses SMILES notation can also be used [5,6]. In the case of nanomaterials even simple mathematical calculations revealing their architecture (similar to the molecular graph) is scarce. That is the reason that, in spite of an increase in the degree of influence of the nanomaterials in modern physical chemistry, industry, and biomedical disciplines, the concept of using QSAR to predict the properties of nanomaterials has not been yet developed.

The aim of the present study is to estimate the ability of a SMILES-like description of nanomaterials as a basis for predicting Young's modulus of these materials. The SMILES-like nomenclature for a given nanomaterial contains data on atom composition and the technological conditions of it's synthesis and is used as basis for calculating optimal descriptors.

However it should be noted that the nomenclature used in the present study is not analogical to the SMILES, since the function of the nomenclature used here nomenclature is restricted to encoding the available information on the genesis of the nanomaterials as commercial products. The SMILES characteristics reflect detailed (2D, 3D, and even quantum chemical) information on molecular architecture.

Data on Young's modulus applied in this study has been taken from [7]. Nanomaterials in ceramic form are included in this data set. The differences between various substances include variations in atomic composition and in temperature of synthesis.

2. Method

The information on nano substances includes the following characteristics: (a) atom composition, (b) type of

* Corresponding author.
E-mail address: aatoropov@yahoo.com (A.A. Toropov).

0009-2614/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.cplett.2006.11.010

substances (bulk or not), and (c) temperature of synthesis. The data for 29 nanomaterials under consideration is presented in Table 1. Each nano structure contains some components which are also included in other nanostructures. The descriptor used for modeling Young's modulus (YM) has been defined as

$$\mathrm{DCW} = \prod_{k=1}^{N} \mathrm{CW}(I_k) \quad (1)$$

where I_k is the component information on the nanostructure (e.g., Al, N, BULK, etc. see Table 1); $\mathrm{CW}(I_k)$ is the correlation weight of the component I_k; and N is the total number of these components in the given nanostructure. A list of all considered components is presented in Table 1.

Thus, the sequence of components applied to a given nanomaterial such as its code and descriptor calculated with Eq. (1) provides an mathematical function of the code.

Using the Monte Carlo method one can calculate the values of the $\mathrm{CW}(I_k)$ that yield correlation coefficients that are as large as possible between Young's modulus (YM) and the DCW for the training set. Having the $\mathrm{CW}(I_k)$ one can then calculate YM by least squares method model:

$$\mathrm{YM} = C_0 + C_1\ \mathrm{DCW} \quad (2)$$

A combination of the $\mathrm{CW}(I_k)$ and Eq. (2) allows the predictive ability of this model concerning nanostructures that are included in the external test set to be estimated.

3. Results

The separation of the considered nanostructures into training and test sets has been done randomly, but according to the following rules: first, all components of the considered species are included in the training set; second, diapasons of Young's modulus values for the training and test sets are approximately the same. Using these rules 29 nanomaterials under consideration have been divided into a training set of 21 nanomaterials and a test set of eight nanomaterials.

The statistical characteristics of the Young's modulus model on the training set and test sets are shown in Table 2. One can see good reproducibility of these characteristics. Also, the standard error of estimating can be considered to be reasonably small. Table 3 contains the numerical values of Young's modulus (experimental and predicted) as well as lists of training and test sets. Table 4 contains the correlation weights obtained in the three probes of the Monte Carlo optimization. Demonstration of the DCW calcula-

Table 1
Information on nanomaterials via codes, defined as the following: BULK denotes that the given nanomaterial is in BULK form (not film); CER denotes that the given nanomaterial is in ceramic form; '%X' is the temperature of synthesis, i.e., %a – 20 °C; %A – 22 °C; %B – 25 °C; %C – 400 °C; %D – 500 °C; %E – 800 °C; %F – 1000 °C; %G – 1100 °C; %H – 1200 °C; %K – 1250 °C; %L – 1400 °C; %M – 1500 °C

ID	Data on genesis of nanomaterial[a] (codes of nanostructures)	Young's modulus (GPa)
1	Al,N,BULK,CER,%B (+)	344.83
2	Al,N,BULK,CER,%F (+)	317.24
3	Al,N,BULK,CER,%L (−)	275.86
4	Al,Al,O,O,O,BULK,CER,%A (+)	376.91
5	Al,Al,O,O,O,BULK,CER,%D (−)	369.92
6	Al,Al,O,O,O,BULK,CER,%E (+)	353.10
7	Al,Al,O,O,O,BULK,CER,%F (−)	329.32
8	Al,Al,O,O,O,BULK,CER,%H (+)	322.23
9	Al,Al,O,O,O,BULK,CER,%K (+)	220.70
10	Al,Al,O,O,O,BULK,CER,%L (+)	225.54
11	Al,Al,O,O,O,BULK,CER,%M (+)	176.65
12	Ti,C,BULK,CER,%A (+)	439.43
13	Ti,C,BULK,CER,%F (+)	344.82
14	Zr,O,O,CER,%A (−)	178.62
15	Zr,O,O,CER,%A (+)	248.28
16	Zr,O,O,CER,%D (−)	137.93
17	Zr,O,O,CER,%E (+)	130.37
18	Zr,O,O,CER,%F (+)	150.01
19	Zr,O,O,CER,%G (+)	210.23
20	Zr,O,O,CER,%H (−)	121.01
21	Zr,O,O,CER,%L (−)	97.931
22	Zr,O,O,CER,%M (+)	88.276
23	Si,C,BULK,CER,%A (+)	410.47
24	Al,Al,Al,Al,Al,Al,O,O,O,O,O,O,O,O,O,Si,Si,O,O,O,O,BULK,CER,%B (+)	127.04
25	Al,Al,Al,Al,Al,Al,O,O,O,O,O,O,O,O,O,Si,Si,O,O,O,O,BULK,CER,%A (−)	143.12
26	Al,Al,Al,Al,Al,Al,O,O,O,O,O,O,O,O,O,Si,Si,O,O,O,O,BULK,CER,%C (+)	130.27
27	Al,Al,Al,Al,Al,Al,O,O,O,O,O,O,O,O,O,Si,Si,O,O,O,O,BULK,CER,%E (+)	102.02
28	Al,Al,Al,Al,Al,Al,O,O,O,O,O,O,O,O,O,Si,Si,O,O,O,O,BULK,CER,%H (+)	27.587
29	Mo,Si,Si,BULK,CER,%A (+)	271.06

[a] Nanomaterials used in the training set are marked by '+'; Nanomaterials used in the test set are marked by '−'.