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 '−'.

Table 2
Statistical characteristics of Young modulus predicted for the training and test sets over three runs of the Monte Carlo optimization

No.	C_0	C_1	Training set, $n = 21$			Test set, $n = 8$			
			r^2	S (GPa)	F	r^2	r^2_{pred}	S (GPa)	F
1	-3720.097 ± 39.863	3945.039 ± 39.215	0.9757	18.25	761	0.8952	0.8880	34.69	51
2	-2182.466 ± 24.518	2346.138 ± 23.232	0.9762	18.03	780	0.8975	0.8762	36.46	53
3	-1829.782 ± 20.808	2155.203 ± 21.109	0.9765	17.93	790	0.9072	0.8804	35.84	59

Table 3
Experimental and calculated values where YM = −3720.235 + 3945.175 DCW of the Young's modulus (YM) for the training and test sets

Titles of nanomaterials according to Ref. [7]	ID	DCW	YM experiment (GPa)	YM calculation (GPa)	YM experiment − calculation (GPa)
Training set					
Aluminum nitride (AlN), bulk	1	1.03350	344.83	357.12017	−12.29017
Aluminum nitride (AlN), bulk	2	1.02081	317.24	307.03297	10.20703
Aluminum oxide (Al2O3),bulk	4	1.04087	376.91[a]	386.17983	−9.26983
Aluminum oxide (Al2O3), bulk	6	1.02734	353.10	332.79176	20.30824
Aluminum oxide (Al2O3), bulk	8	1.01882	322.23[a]	299.18831	23.04169
Aluminum oxide (Al2O3), bulk	9	0.99897	220.70	220.88267	−0.18267
Aluminum oxide (Al2O3), bulk	10	1.00061	225.54	227.34564	−1.80564
Aluminum oxide (Al2O3), bulk	11	0.99668	176.65	211.82746	−35.17746
Titanium carbide (TiC), bulk	12	1.05123	439.43	427.05857	12.37143
Titanium carbide (TiC), bulk	13	1.03211	344.82[a]	351.60433	−6.78433
Zirconium oxide (ZrO2)	15	1.00539	248.28	246.21894	2.06106
Zirconium oxide (ZrO2)	17	0.98593	130.37	169.43313	−39.06313
Zirconium oxide (ZrO2)	18	0.98074	150.01[a]	148.97011	1.03989
Zirconium oxide (ZrO2)	19	0.99582	210.23	208.46891	1.76109
Zirconium oxide (ZrO2)	22	0.95650	88.276	53.34434	34.93166
Silicon carbide (SiC), bulk	23	1.04781	410.47	413.54853	−3.07853
Mullite (3Al2O3 2SiO2), bulk	24	0.97171	127.04	113.33305	13.70695
Mullite (3Al2O3 2SiO2), bulk	26	0.97577	130.27	129.35619	0.91381
Mullite (3Al2O3 2SiO2), bulk	27	0.96485	102.02	86.26609	15.75391
Mullite (3Al2O3 2SiO2), bulk	28	0.95685	27.587	54.70657	−27.11957
Molybdenum Silicide (MoSi2), bulk	29	1.01203	271.06	272.38543	−1.32543
Test set					
Aluminum nitride (AlN), bulk	3	0.99951	275.86	223.00185	52.85815
Aluminum oxide (Al2O3),bulk[b]	5	1.02876	369.92[a]	338.39795	31.52205
Aluminum oxide (Al2O3), bulk	7	1.02193	329.32[a]	311.46933	17.85067
Zirconium oxide (ZrO2)	14	0.99892	178.62[a]	220.66937	−42.04937
Zirconium oxide (ZrO2)[b]	16	0.98729	137.93	174.81336	−36.88336
Zirconium oxide (ZrO2)	20	0.97776	121.01[a]	137.18408	−16.17408
Zirconium oxide (ZrO2)	21	0.96028	97.931	68.23705	29.69395
Mullite (3Al2O3 2SiO2), bulk	25	0.97756	143.12	136.40682	6.71318

[a] In the case of interval values of Young's modulus ($x \ldots y$) the average $(x + y)/2$ has been used.
[b] If a I_k is absent in the training set, then CW(I_k) is defined as being equal to 1.0.

tion is presented in Table 5. Graphical illustrations of correlation between experimental and calculated values of the Young's modulus values are shown in Figs. 1 and 2, respectively.

On the basis of the analyzed data obtained by three runs of the Monte Carlo optimization there are three categories of the components of the nanomaterials: the first category includes components with values of correlation weights more than 1.0 over all three probes (C, Al, BULK, %a, %A, %C, %B, %G); the second category includes components with various values of the correlation weights, i.e., both more than 1.0 and less than 1.0 (O, Si, Ti, Mo, %E); and the third category includes components with correlation weights less than 1.0 (N, Zr, %F, %H, %K, %L, %M). The first category can be characterized as promoters of the increase of the Young's modulus value. The third category can be characterized as promoters of the decrease of the Young's modulus value. Finally the second category can be interpreted as components of an undefined role. The separation of components into these categories could help in heuristic searching for components of nanomaterials with preferable values of Young's modulus.

In addition four random divisions into training and test sets have been also studied. These divisions are shown in Table 6. The statistical characteristics of these models are listed in Table 7. One can see from Table 7, that these models have reasonable good statistical quality. It is to be noted that some attributes of nanomaterials are absent in the training for the first, third, and fourth divisions. The correlation weights of attributes in such cases are equal to 1.0.

Table 4
Correlation weights of component information on nanomaterials for three runs of the Monte Carlo optimization

I_k	CW(I_k) in run 1	CW(I_k) in run 2	CW(I_k) in run 3
Zr	0.9606355	0.9725848	0.9857441
Ti	0.9948871	0.9879302	1.0194368
Si	0.9916462	0.9788797	1.0103312
O	0.9993185	1.0022752	1.0011168
Mo	0.9844970	1.0017344	0.9753210
N	0.9900020	0.9910674	0.9839821
C	1.0078754	1.0286025	1.0297504
Al	1.0017618	1.0068783	1.0440719
Bulk	1.0068267	1.0456045	1.0393599
CER	1.0411087	1.0475217	1.0085239
,	0.9971167	0.9931635	0.9799608
%a	1.0183324	1.0398939	1.0499022
%M	0.9688145	0.9567842	0.9509630
%L	0.9726380	0.9631401	0.9585375
%K	0.9710456	0.9613364	0.9561908
%H	0.9903392	0.9914617	0.9910316
%G	1.0086406	1.0240313	1.0300265
%F	0.9933651	0.9978241	0.9986069
%E	0.9986187	1.0060672	1.0080645
%C	1.0099232	1.0252212	1.0303073
%B	1.0057196	1.0177154	1.0212428
%A	1.0117729	1.0291263	1.0337684

Table 5
Example calculations of DCW by Eq. (1) for nanomaterial with ID 1 (Table 1)

No.	I_k	CW(I_k) in run 1
1	Al	1.00176
2	,	0.99712
3	N	0.99000
4	,	0.99712
5	Bulk	1.00683
6	,	0.99712
7	CER	1.04111
8	,	0.99712
9	%B	1.00572

DCW = 1.03350.

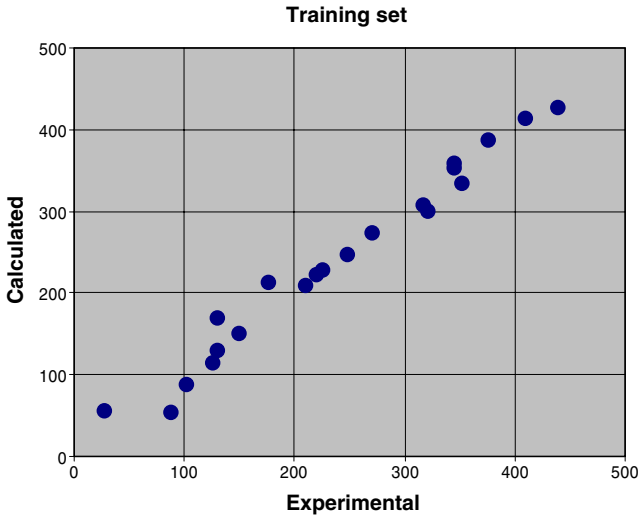

Fig. 1. Correlation between experimental and calculated Young's modulus (in GPa) values for the training set.

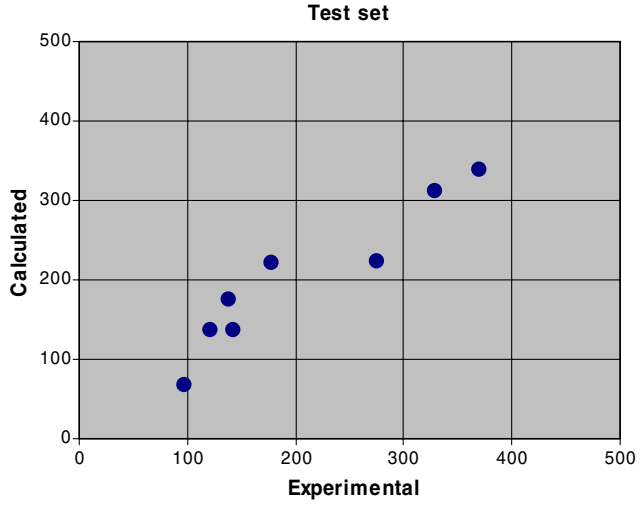

Fig. 2. Correlation between experimental and calculated Young's modulus (in GPa) values for the test set.

Table 6
Added divisions into training and test sets

Division number	List of external test sets
1	5, 7, 10, 14, 16, 21, 23, 25
2	3, 5, 7, 10, 12, 17, 25, 27
3	1, 5, 7, 14, 16, 20, 21, 25
4	1, 3, 5, 7, 16, 21, 25, 28

In fact information concerning nanomaterials in view of the data displayed in Table 1 corresponds to instruction on how to carry out the synthesis of a given substance. Under such circumstances the suggested approach can be used as a tool of estimation of the Young's modulus value for substances that can be produced under technological conditions which have not been used before. In other words, the suggested approach can help to save time, money and material resources in searching for nanomaterials with appropriate values of Young's modulus. Most probably the accuracy of the prediction will increase with increasing number of nanomaterials used in the training set.

In recent years efforts to define nanostructures in rational mathematical terms have been carried out [8]. If these efforts will result in systematic representations of the architecture of nanomaterials, then such representations will enable descriptors to be defined which are similar to the existing topological (2D), geometric (3D), and quantum chemical descriptors calculated with 'classical' molecular graphs. This situation will complement the approach described in this article, the systematic rational representation of data concerning nanomaterials can be used to facilitate the process of developing and expanding on the concept of optimal descriptors.

Table 7
Statistical characteristics of the four models with other splits into training and test sets

Division	C_0	C_1	r^2	s	F	r^2	r^2_{pred}	S	F
1	-10438.04 ± 110.28	10272.79 ± 105.82	0.9715	19.11	649	0.9210	0.8947	38.37	70
2	-423.368 ± 6.388	556.536 ± 5.041	0.9822	14.75	1047	0.9155	0.8926	40.22	65
3	-2117.217 ± 22.072	2362.637 ± 21.538	0.9763	17.68	782	0.8625	0.8496	43.71	38
4	-1890.456 ± 23.684	2069.096 ± 22.579	0.9768	16.53	799	0.9253	0.9177	37.17	74

4. Conclusions

Twenty-nine different nanomaterials characterized by by experimental studies were selected to test a new theoretical methodology. The applied approach allows for the prediction of reasonable numerical values of the Young's modulus of eight randomly selected nanomaterials for a test set from a training set of 21 nanomaterials.

Acknowledgement

The authors thank for support the High Performance Computational Design of Novel Materials (HPCDNM) Project funded by the US Department of Defense through the US Army Engineer Research and Development Center (Vicksburg, MS) Contract #W912HZ-06-C-0057.

References

[1] A.A. Toropov, A.P. Toropova, Russ. J. Coord. Chem. 24 (1998) 81.
[2] D. Amic, D. Beslo, B. Lucic, S. Nikolic, N. Trinajstic, J. Chem. Inf. Comput. Sci. 38 (1998) 819.
[3] M. Randic, S.C. Basak, J. Chem. Inf. Comput. Sci. 39 (1999) 261.
[4] M. Randic, M. Pompe, J. Chem. Inf. Comput. Sci. 41 (2001) 631.
[5] D. Vidal, M. Thormann, M. Pons, J. Chem. Inf. Model. 45 (2005) 386.
[6] A.A. Toropov, A.P. Toropova, D.V. Mukhamedzhanova, I. Gutman, Indian J. Chem. 44A (2005) 1545.
[7] J.F. Shackelford, W. Alexander, The CRC Materials Science and Engineering Handbook, third ed., CRC press, Boca Raton FL 33431, 2000, p. 1980.
[8] T.C. Dinadayalane, J. Leszczynski, Proc. Constr. Mater. Sci. Machinery Constr. 36 (2006) 58.