Int. J. Environ. Res., 9(4):1211-1216, Autumn 2015
ISSN: 1735-6865

Application of the Monte Carlo Method to Prediction of Dispersibility of Graphene in Various Solvents

Toropova, A.P.[1], Toropov, A.A.[1*], Veselinovic, J.B.[2], Vesélinovic, A.M.[2] ,Benfenati, E.[1], Leszczynska, D.[3] and Leszczynski, J.[4]

[1]IRCCS-Istituto di Ricerche Farmacologiche Mario Negri, Via La Masa 19, Milano, Italy

[2]University of Niš, Faculty of Medicine, Department of Chemistry, Niš, Serbia

[3] Interdisciplinary Nanotoxicity Center, Department of Civil and Environmental Engineering, Jackson State University, 1325 Lynch St, Jackson, MS 39217-0510, USA

[4] Interdisciplinary Nanotoxicity Center, Department of Chemistry and Biochemistry, Jackson State University, 1400 J. R. Lynch Street, P.O. Box 17910, Jackson, MS 39217, USA

Received 17 Nov. 2014; Revised 22 Jan. 2015; Accepted 25 Jan. 2015

ABSTRACT: The dispersibility of graphene is modeled as a mathematical function of the molecular structure of solvent represented by simplified molecular input-line entry systems (SMILES) together with the graph of atomic orbitals (GAO). The GAO is molecular graph where atomic orbitals e.g. 1s1, 2p4, 3d7 etc., are vertexes of the graph instead of the chemical elements used as the graph vertexes in the traditionally used molecular graph (hydrogen suppressed molecular graph or hydrogen filled molecular graph). The optimal descriptors calculated with the Monte Carlo method were used to build up one variable correlations "descriptor- dispersibility". The CORAL software is used as a tool to build up the model. Based on the results of calculations the structural features which are promoters of increase or those which are promoters of decrease of the dispersibility are detected and discussed. The predictive potential of the used approach is checked up with three random and non identical splits of available data into the training, calibration, and validation (invisible during building up the model) sets. The statistics for external validation sets are the following: n=11, r2=0.6379, s=0.392 (split 1); n=8, r^2=0.7308, s=0.378 (split 2); and n=5, r^2=0.7797, s=0.504 (split 3).

Key words: QSPR, Monte Carlo method, graphene,Dispersibility,CORAL software

INTRODUCTION

One of the recent, most spectacular breakthroughs in nanotechnology are associated with various applications of graphene. This is due to its distinctive characteristics - graphene offers a unique combination of electrical, optical, thermal, and mechanical properties (Yousefinejad and Hemmateenejad, 2014).After being a subject of numerous studies in scientific laboratories graphene has been fast transfer to the manufacturing plants. In order to fully utilize its potential more basic information about this unique species are necessary.

Though there is continuous progress in theoretical studies, many aspects of graphene are still not well understood. Theoretical works that provide an insight on interactions in systems "graphene-solvent" are very rare. Therefore, there is a considerable demand for prediction of influences of solvents for its dispersibility. Quantitative structure - property relationships (QSPRs) provide an approach that could used to solve this task (Afantitis et al., 2011; Furtula and Gutman , 2011; Furtula et al., 2013; García *et al.*, 2011; Garro Martinez *et al.*, 2011; Nesmerak *et al.*, 2013; Roy and Paul, 2009; Toropov and Toropova , 2003). The aim of the present study is the evaluation of ability of the hybrid optimal descriptors calculated with simplified molecular input-line entry system (SMILES) together with graph of atomic orbitals (GAO) (Toropov and Toropova , 2003) to provide an efficient tool to build up model of dispersibility of graphene in various solvents.

*Corresponding author E-mail: andrey.toropoc@marionegri.it

MATERIAL & METHODS

The numerical data on dispersibility of graphene (D_G, μg/ml) in different solvents (n=40) is taken from

the literature (Hernandez *et al.*, 2010). The decimal logarithm logDG is examined as the endpoint. Table 1 contains the molecular structures and the numerical data on the graphene dispersibility together with SMILES prepared by ACD/Chemsketh software (www.acdlabs.com). Three random different splits of the experimental data into the visible training and calibration sets and invisible validation set are examined as the basis for the computational experiments.
Optimal descriptor used in this work is calculated as the following:

$$DCW(SMILES, GAO, Threshold, N_{epoch}) = \Sigma CW(Sk) + \Sigma CW(AOj) \quad (1)$$

where Threshold (Toropova *et al.*, 2011) is coefficient for classification of various molecular features extracted from SMILES and / or GAO into two classes: (i) active (in this case correlation weight is involved in the modeling process); and (ii) rare (in this case correlation weight is not involved in the modeling process); the Nepoch (Toropova *et al.*, 2011) is the number of epochs of the Monte Carlo optimization which gives the best statistical quality for the calibration set; Sk is a fragment of SMILES notation i.e. one or two symbols from SMILES (e.g. 'Cl', 'Br', etc. cannot be examined separately); and AOj is the vertex degree in GAO (Toropov and Toropova 2003). Having numerical data on correlation weights which give the preferable statistics for the calibration set, one can calculate (using the training set) the model

$$logDG = C0 + C1*DCW(SMILES, GAO, Threshold, N_{epoch}) \quad (2)$$

The predictive potential of the model should be tested with external validation set which is invisible for build up the model.

RESULTS & DISCUSSION

The statistical quality of the suggested models is the following:
Split 1

$$logD_G = 0.4468 (\pm 0.0102) + 0.0303(\pm 0.0024) * DCW(SMILES, GAO, 1,20) \quad (3)$$

n=21, r^2=0.7302, q^2=0.5430, s=0.181, F=51 (training set)
n=8, r^2=0.8254, s=0.200 (calibration set)
n=11, r^2=0.6379, s=0.392 (validation set)
Split 2

$$logD_G = 0.3032(\pm 0.0147) + 0.0363 (\pm 0.0022) * DCW(SMILES, GAO,1,9) \quad (4)$$

n=25, r^2=0.6832 q^2=0.5463, s=0.182, F=50 (training set)
n=7, r^2=0.8457, s=0.182 (calibration set)
n=8, r^2=0.7308, s=0.378 (validation set)
Split 3

$$logD_G = 0.3138 (\pm 0.0090) + 0.0229 (\pm 0.0010) * DCW(SMILES, GAO, 1,15) \quad (5)$$

n=28, r^2=0.6972 q^2=0.6054, s=0.149, F=60 (training set)
n=7, r^2=0.8358, s=0.314 (calibration set)
n=5, r^2=0.7797, s=0.504 (validation set)

The statistical characteristics of the models for the grapheme dispersibility in the same solvents calculated by the multiple linear regression analysis with involving of topological, geometrical, and quantum chemical descriptors are the following (Yousefinejad and Hemmateenejad, 2014): (i) minimal r^2=0.47 and maximal r^2=0.913; and (ii) standard error of estimation: minimum is equal to 0.172 while the maximum is equal to 0.330. Thus, one can conclude that the statistical quality of models calculated with Eqs. 3-5 is comparable with the models described in work (Yousefinejad and Hemmateenejad, 2014).

The current study provides additional important information. The performed analysis for a group of runs of the Monte Carlo optimization indicates that there are stable promoters (Toropov *et al.*, 2011) of logDG increase. The list of promoters includes: (i) presence of nitrogen; (ii) branching; (iii) presence of rings; (iv) presence of triple covalent bonds; and (v) presence of vertex degree equal to 9 for 2p3. In addition, we are also able to identify the promoters of logDG decrease. These are the following descriptors: (i) presence of aromatic systems; (ii) presence of vertex degree equal to 6 for 2p2; and (iii) presence of vertex degree equal to 9 for 2s2. This proofs that the models calculated with Eqs. 3-5 possess mechanistic interpretations (OECD, 2007).

One also notices that the optimal descriptors calculated with solely SMILES or with solely GAO approach provide poorer models for the endpoint than the hybrid technique. This indicates a promising potential of the hybrid descriptor (Toropova *et al.*, 2012) as a tool for the QSPR/QSAR analyses.

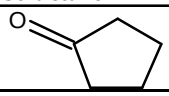
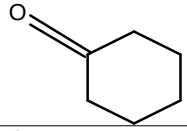
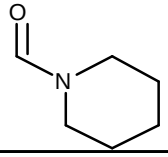
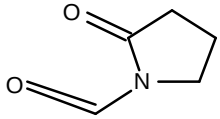
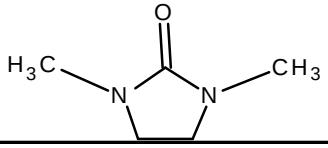
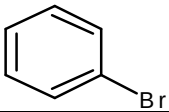
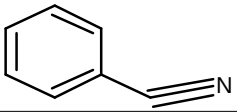
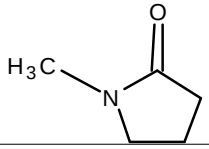
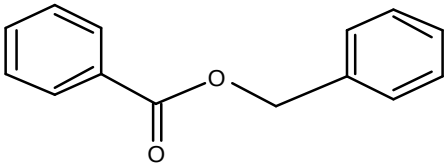
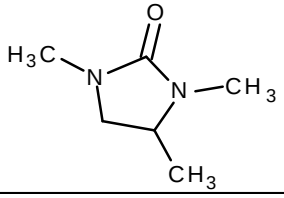
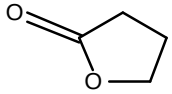
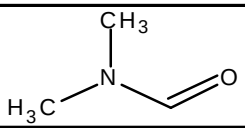
CONCLUSIONS

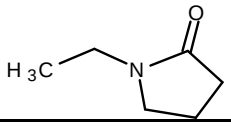
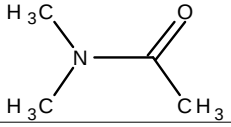
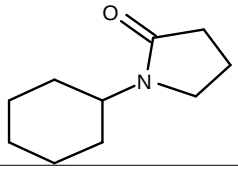
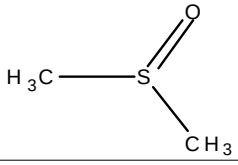
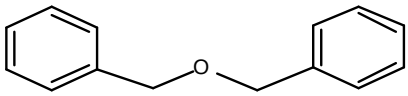
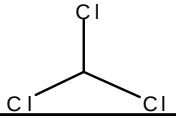
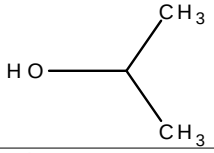
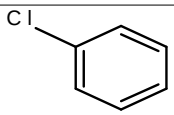
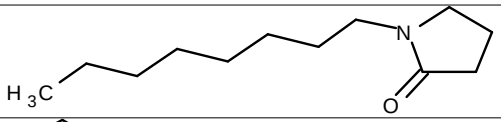
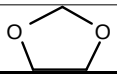
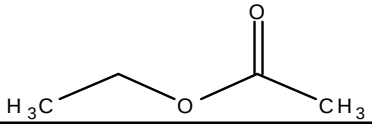
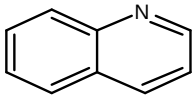
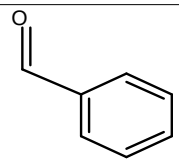
A QSPR approach was used for analysis of the graphene dispersibility. The tested here descriptors calculated with the CORAL software (http://www.insilico.eu/coral) are demonstrated as capable components that can be used as a tool to build up efficient QSPR for graphene dispersibility. The statistical quality of developed models is influenced by details of splits of the experimental data into visible training and calibration sets, and invisible validation set. In addition, the applied approach possesses a mechanistic interpretation. The developed here models obey to OECD principles (OECD, 2007).

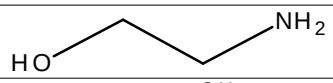
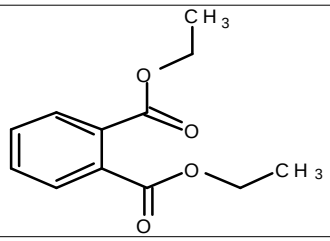
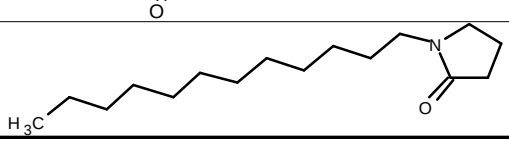
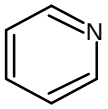
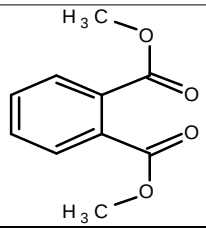
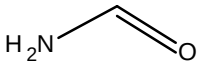
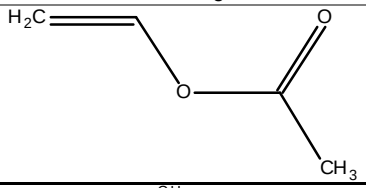
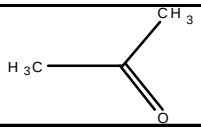
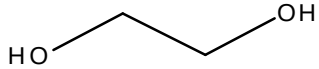
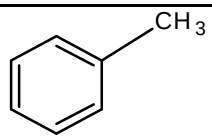
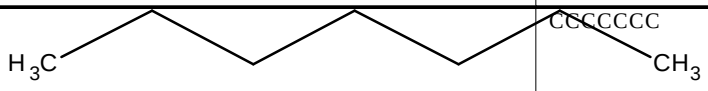
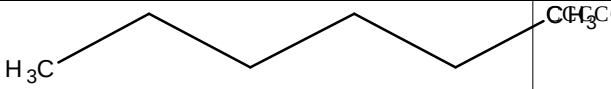
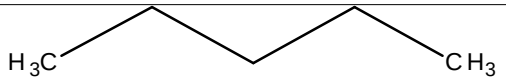
ACKNOWLEDGEMENTS

The authors acknowledge support from the EC

Table 1.Structures and SMILES of solvents and the numerical data on the graphene dispersibility

ID	Structure	SMILES	$\log D_G$
1		O=C1CCCC1	0.93
2		O=C1CCCCC1	0.86
3		O=CN1CCCCC1	0.86
4		O=C1CCCN1C=O	0.74
5		O=C1N(C)CCN1C	0.73
6		Brc1ccccc1	0.71
7		N#Cc1ccccc1	0.68
8		O=C1CCCN1C	0.67
9		O=C(OCc1ccccc1)c2ccccc2	0.67
10		O=C1N(C)CC(C)N1C	0.66
11		O=C1CCCO1	0.61
12		CN(C)C=O	0.61

13		O=C1CCCN1CC	0.60
14		CN(C)C(C)=O	0.59
15		O=C2CCCN2C1CCCCC1	0.57
16		CS(C)=O	0.57
17		C(OCc1ccccc1)c2ccccc2	0.54
18		ClC(Cl)Cl	0.53
19		CC(C)O	0.49
20		Clc1ccccc1	0.46
21		O=C1CCCN1CCCCCCCCC	0.45
22		C1COCO1	0.45
23		CC(=O)OCC	0.41
24		c1cccc2cccnc12	0.41
25		O=Cc1ccccc1	0.40

26		NCCO	0.40
27		O=C(OCC)c1ccccc1C(=O)OCC	0.34
28		O=C1CCCN1CCCCCCCCCCCC	0.32
29		c1ccccn1	0.3
30		COC(=O)c1ccccc1C(=O)OC	0.26
31		NC=O	0.23
32		CCO	0.20
33		CC(=O)OC=C	0.18
34		CC(C)=O	0.08
35	H_2O	O	0.04
36		OCCO	0.00
37		Cc1ccccc1	-0.10
38		CCCCCCC	-0.52
39		CCCCCC	-0.70
40		CCCCC	-0.80

project NANOPUZZLES (Project Reference: 309837), EU FP7 project PreNanoTox (contract 309666), and EU project PROSIL funded under the LIFE program (project LIFE12 ENV/IT/000154). D.L. and J.L. acknowledge support from the National Science Foundation (CREST HRD-0833178 and EPSCoR 362492-190200-01\NSFEPS-0903787).

REFERENCES

Afantitis, A., Melagraki, G., Koutentis, P.A., Sarimveis, H. and Kollias, G. (2011). Ligand - based virtual screening procedure for the prediction and the identification of novel b-amyloid aggregation inhibitors using Kohonen maps and Counterpropagation Artificial Neural Networks. European Journal of Medicinal Chemistry, **46(2),** 497-508.

Furtula, B. and Gutman, I. (2011). Relation between second and third geometric-arithmetic indices of trees. Journal of Chemometrics, **25(2),** 87-91.

Furtula, B., Gutman, I. and Dehmer, M. (2013). On structure-sensitivity of degree-based topological indices. Applied Mathematics and Computation, **219,** 8973-8978.

García, J., Duchowicz, P.R., Rozas, M.F., Caram, J.A., Mirífico, M.V., Fernández, F.M. and Castro, E.A. (2011). A comparative QSAR on 1,2,5-thiadiazolidin-3-one 1,1-dioxide compounds as selective inhibitors of human serine proteinases. Journal of Molecular Graphics and Modelling, **31,**10-19.

Garro Martinez, J.C., Duchowicz, P.R., Estrada, M.R., Zamarbide, G.N. and Castro, E.A. (2011). QSAR Study and Molecular Design of Open-Chain Enaminones as Anticonvulsant Agents. International Journal of Molecular Sciences, **12(12),** 9354-9368.

Hernandez, Y., Lotya, M., Rickard, D., Bergin, D.S. and Jonathan N. Coleman, J. N. (2010). Measurement of multicomponent solubility parameters for graphene facilitates solvent discovery. Langmuir, **26(5),** 3208-3213.

Nesmerak, K., Toropov, A.A., Toropova, A.P., Kohoutova, P. and Waisser, K. (2013). SMILES-based quantitative structure-property relationships for half-wave potential of N-benzylsalicylthioamides. European Journal of Medicinal Chemistry, **67,** 111-114.

Organization for Economic Co-operation and Development (OECD) (2007). Guidance Document on The Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models, No. 69. http://www.oecd.org/dataoecd/55/35/38130292.pdf

Roy, K. and Paul, S. (2009). Exploring 2D and 3D QSARs of 2, 4-diphenyl-1, 3-oxazolines for ovicidal activity against Tetranychus urticae. QSAR & Combinatorial Science, **28(4),** 406-425.

Toropov, A.A. and Toropova, A.P. (2003). QSPR modeling of alkanes properties based on graph of atomic orbitals. Journal of Molecular Structure: THEOCHEM, **637(1),** 1-10.

Toropov, A.A., Toropova, A.P., Benfenati, E., Gini, G., Leszczynska, D. and Leszczynski, J. (2011). SMILES-based QSAR approaches for carcinogenicity and anticancer activity: Comparison of correlation weights for identical SMILES attributes. Anti-Cancer Agents in Medicinal Chemistry, **11(10),** 974-982.

Toropova, A.P., Toropov, A.A., Benfenati, E., Gini, G., Leszczynska, D. and Leszczynski, J. (2011). CORAL: Quantitative Structure-Activity Relationship Models for Estimating Toxicity of Organic Compounds in Rats. Journal of Computational Chemistry, **32(12),** 2727-2733.

Toropova, A.P., Toropov, A.A., Martyanov, S.E., Benfenati, E., Gini, G., Leszczynska, D. and Leszczynski, J. (2012). CORAL: QSAR modeling of toxicity of organic chemicals towards Daphnia magna. Chemometrics and Intelligent Laboratory Systems, **110,** 177-181.

Yousefinejad, S. and Hemmateenejad, B. (2014). A chemometrics approach to predict the dispersibility of graphene in various liquid phases using theoretical descriptors and solvent empirical parameters. Colloids and Surfaces A Physicochemical and Engineering Aspects, **441,** 766-775.