J Math Chem (2013) 51:2230–2237
DOI 10.1007/s10910-013-0211-2

ORIGINAL PAPER

Optimal descriptor as a translator of eclectic information into the prediction of thermal conductivity of micro-electro-mechanical systems

Alla P. Toropova · Andrey A. Toropov · Tomasz Puzyn · Emilio Benfenati · Danuta Leszczynska · Jerzy Leszczynski

Received: 14 May 2013 / Accepted: 10 June 2013 / Published online: 20 June 2013

Abstract Micro-electro-mechanical systems (MEMS) are involved in various fields of nanotechnology. MEMS are characterized by complex and unclear molecular architecture. However, in most cases information about chemical composition and conditions of synthesis is available. One-variable models for thermal conductivity of MEMS are suggested. These models are based on the representation of MEMS by their chemical composition and technological attributes. We have examined three random splits of available data into the training set and validation set. The average statistical characteristics of these models are quite good. Development of suggested here models is carried out without information on molecular architecture of MEMS.

Keywords Micro-electro-mechanical system · Nanotechnology · Optimal descriptor · CORAL software

A. P. Toropova · A. A. Toropov (✉) · E. Benfenati
IRCCS, Istituto di Ricerche Farmacologiche Mario Negri, Via La Masa 19, 20156 Milan, Italy
e-mail: aatoropov@yahoo.com

T. Puzyn
Laboratory of Environmental Chemometrics, Faculty of Chemistry, University of Gdansk, ul. Sobieskiego 18/19, 80-952 Gdańsk, Poland

D. Leszczynska
Interdisciplinary Nanotoxicity Center, Department of Civil and Environmental Engineering, Jackson State University, 1325 Lynch St, Jackson, MS 39217-0510, USA

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

1 Introduction

There are various ways to predict properties of molecular systems. Among them are approaches linking computational and experimental studies. Such techniques are represented by quantitative structure—property/activity relationships (QSPRs/QSARs). The majority of QSPRs/QSARs studies are based on representation of substances by information on their molecular architecture [1–7]. The exception are QSPR/QSAR models based on physicochemical parameters such as, octanol / water partition coefficient, water solubility, and others. However, also substances with unclear or extremely large and complex molecular architecture do exist [8–11]. For example, among them are various allotropic forms of inorganic substances [9], polymers [10], peptides [11], and various nano substances [12].

Obviously in the case of substances with unclear molecular architecture, the standard QSPR/QSAR approach cannot be applied. In such systems the QSPR/QSAR models can be built up using alternative information, e.g. (i) the above-mentioned available physicochemical parameters, such as octanol water partition coefficient, solubility [13]; and interestingly, by (ii) technological conditions and parameters selected for manufacturing of various target substances [14].

Micro-electro-mechanical systems (MEMS) [15] represent a technology that in its most general form can be defined as miniaturization of mechanical and electro-mechanical elements. Nanotechnology provides the ability to manipulate substances at the atomic or molecular level to make something useful at the nano-dimensional scale. Many devices and structures of nanotechnology are made using MEMS techniques.

The aim of the present study is estimation of optimal descriptors to be applied as a tool of prediction of thermal conductivity of MEMS which were not examined in the experiment.

1.1 Method

1.2 Data

Experimental data on the thermal conductivity of MEMS taken in [15]. The logarithm of the thermal conductivity expressed in [W/m/K] has been used as the endpoint. The collection of MEMS has been selected according to two principles (i) these MEMS have (partially) the same technological attributes (Table 1); and (ii) their number is as large as possible (Table 2). These substances were three times randomly split into the sub-training set (i.e. a group of MEMS which are "producer of model"), calibration set (i.e. a group of MEMS which are "critic of model"), test set (i.e. a group of MEMS which are "preliminary estimator of model"), and validation set (i.e. a group of MEMS which are "final estimator of model").

1.3 Optimal descriptors

Optimal descriptors for MEMS are calculated as the following

$$\mathbf{DCW}\left(\mathbf{T}, \mathbf{N_{epoch}}\right) = \sum \mathbf{CW}\left(\mathbf{Code_k}\right) \tag{1}$$

Table 1 Technological attributes and their codes, which are using for building up model of thermal conductivity for MEMS

Temperature (C)	Code of the temperature
20	%1
25	%1
27	%1
80	%1
100	%1
127	%1
150	%1
200	%2
250	%2
273.1	%2
315	%3
350	%3
400	%4
425	%4
500	%5
540	%5
600	%6
650	%6
700	%7
800	%8
875	%9
1,000	%10
1,100	%11
1,200	%12
1,250	%12
1,327	%13
1,400	%14
1,530	%15
1,600	%16
2,300	%17
Status of MEMS	**Code of status**
Ceramic	1
Single crystal	2
Cubic	3
CVD	4
Glass	5

CVD chemical vapor deposition

where $CW(Code_k)$ are correlation weights for technological codes related to a MEMS (Table 1). The T is threshold. The threshold is the coefficient for classification of codes into two classes rare (noise) and not rare. If the frequency of a code in the sub-training set is less than the Threshold, the correlation weight for this code will be zero and consequently the code will be not involved in building up model. The number

Table 2 MEMS, their codes and data on the decimal logarithm of thermal conductivity, as well three splits of available data into the sub-training set (+), calibration set (−), test set (#), and validation set (*)

MEMS	Split			Codes for MEMS	lgTC
	1	2	3		
AlN-1	+	+	+	Al.N.%6	1.302
AlN-2	*	−	*	Al.N.%4	1.345
AlN-3	−	+	+	Al.N.%1	1.479
Al_2O_3-1	#	−	−	Al.Al.O.O.O.2	1.699
Al_2O_3-2	+	+	*	Al.Al.O.O.O.1.%14	0.735
Al_2O_3-3	−	−	+	Al.Al.O.O.O.1.%1	1.399
Al_2O_3-4	−	*	#	Al.Al.O.O.O.1.%3	1.189
Al_2O_3-5	*	*	+	Al.Al.O.O.O.1.%5	1.165
Al_2O_3-6	+	#	#	Al.Al.O.O.O.2.%1	1.634
Al_2O_3-7	*	#	+	Al.Al.O.O.O.2.%3	1.293
Al_2O_3-8	*	+	−	Al.Al.O.O.O.2.%8	1.084
BN-1	*	#	*	B.N.1.%3	1.458
BN-2	−	#	*	B.N.1.%7	1.431
BN-3	*	−	+	B.N.1.%10	1.425
Cd	−	+	−	Cd.%1	1.986
Cr	+	−	+	Cr.%1	1.956
CrB_2	−	#	−	Cr.B.B.%1	1.311
Cr_3C_3	−	+	−	Cr.Cr.Cr.C.C.C.1	2.278
GaAs	+	+	+	Ga.As.%1	1.663
Mo	−	+	*	Mo.%1	2.140
$MoSi_2$-1	#	*	−	Mo.Si.Si.%1.1	1.732
$MoSi_2$-2	#	+	#	Mo.Si.Si.%4.1	1.490
$MoSi_2$-3	#	−	#	Mo.Si.Si.%5.1	1.345
$MoSi_2$-4	−	#	+	Mo.Si.Si.%6.1	1.377
$MoSi_2$-5	#	−	−	Mo.Si.Si.%9.1	1.284
$MoSi_2$-6	+	*	#	Mo.Si.Si.%11.1	1.234
$(Al_2O_3)_3$*$(SiO_2)_2$-1	+	*	+	Al.Al.O.O.O.Al.Al.O.O.O.Al.Al.O.O.O.Si.O.O.Si.O.O.%1.1	0.782
$(Al_2O_3)_3$*$(SiO_2)_2$-2	#	#	#	Al.Al.O.O.O.Al.Al.O.O.O.Al.Al.O.O.O.Si.O.O.Si.O.O.%2.1	0.735
$(Al_2O_3)_3$*$(SiO_2)_2$-3	−	*	*	Al.Al.O.O.O.Al.Al.O.O.O.Al.Al.O.O.O.Si.O.O.Si.O.O.%4.1	0.663
$(Al_2O_3)_3$*$(SiO_2)_2$-4	+	#	−	Al.Al.O.O.O.Al.Al.O.O.O.Al.Al.O.O.O.Si.O.O.Si.O.O.%6.1	0.621
$(Al_2O_3)_3$*$(SiO_2)_2$-5	*	−	+	Al.Al.O.O.O.Al.Al.O.O.O.Al.Al.O.O.O.Si.O.O.Si.O.O.%8.1	0.599
$(Al_2O_3)_3$*$(SiO_2)_2$-6	#	*	*	Al.Al.O.O.O.Al.Al.O.O.O.Al.Al.O.O.O.Si.O.O.Si.O.O.%10.1	0.575
$(Al_2O_3)_3$*$(SiO_2)_2$-7	*	#	*	Al.Al.O.O.O.Al.Al.O.O.O.Al.Al.O.O.O.Si.O.O.Si.O.O.%12.1	0.575
$(Al_2O_3)_3$*$(SiO_2)_2$-8	−	+	*	Al.Al.O.O.O.Al.Al.O.O.O.Al.Al.O.O.O.Si.O.O.Si.O.O.%14.1	0.575

Table 2 continued

MEMS	Split			Codes for MEMS	lgTC
	1	2	3		
Ni	*	+	#	Ni.%1	1.957
Pt	#	+	–	Pt.%1	1.863
SiC-1	*	–	+	Si.C.3.4.%1	2.082
SiC-2	+	#	#	Si.C.3.4.%6	1.319
SiC-3	–	#	*	Si.C.3.4.%8	1.407
SiC-4	–	–	#	Si.C.3.4.%10	1.329
SiC-5	#	–	#	Si.C.3.4.%13	1.539
SiO_2-1	–	*	*	Si.O.O.1.%2	0.017
SiO_2-2	+	*	–	Si.O.O.1.%4	0.097
SiO_2-3	#	–	#	Si.O.O.1.%8	0.223
SiO_2-4	#	#	–	Si.O.O.1.%12	0.320
SiO_2-5	#	+	*	Si.O.O.1.%16	0.400
SiO_2-6	*	–	+	Si.O.O.5.%1	0.140
SiO_2-7	*	*	–	Si.O.O.5.%2	0.107
SiO_2-8	+	#	–	Si.O.O.5.%3	0.134
SiO_2-9	#	*	#	Si.O.O.5.%4	0.176
SiO_2-10	*	#	+	Si.O.O.5.%7	0.255

of epochs (N_{epoch}) is the number of cycles (sequence of modification of correlation weights for all codes involved in building up model) for the Monte Carlo optimization of correlation weights. The optimal correlation weights should give the maximum for the target function described in [16]. The predictive potential of this model should be checked up with external validation set. Preferable values of these parameters were defined empirically as described in the literature [17]. The space of searching to define preferable the number of epochs (N_{epoch}) and the preferable threshold (T) is the following: the range of N_{epoch} from 1 to 100; and the range of T from 1 to 5. The calculation of the above-mentioned optimal descriptors has been carried out with the CORAL software [16]. Table 3 contains correlation weights of codes which are used to calculate optimal descriptors for split 1, 2, and 3.

2 Results and discussion

The models of decimal logarithm of thermal conductivity (lgTC) are the following:

$$\text{LgTC} = 0.7914\,(\pm 0.009) + 0.1556\,(\pm 0.0030) * \text{DCW}\,(1, 44) \quad (2)$$

$n = 11, r^2 = 0.9752, q^2 = 0.9628,\ s = 0.097,\ F = 355$ (sub-training set)

$n = 14,\ r^2 = 0.7977,\ s = 0.373$ (calibration set)

$n = 13,\ r^2 = 0.9601,\ s = 0.419$ (test set)

$n = 13,\ r^2 = 0.8229,\ s = 0.351$ (validation set)

Table 3 Lists of codes and their correlation weights calculated by the Monte Carlo method for three random splits of available data into training set (it includes sub-training, calibration, and test set) and external validation set

Split 1		Split 2		Split 3	
Code	CW (code)	Code	CW (code)	Code	CW (code)
%1	3.97200	%1	2.84600	%1	1.52100
%10	0.0	%10	1.29900	%10	0.0
%11	2.57800	%11	1.64050	%11	0.0
%12	0.0	%12	0.21350	%12	0.0
%13	0.0	%13	0.0	%13	0.0
%14	−0.21050	%14	0.0	%2	0.0
%16	0.0	%2	0.0	%3	0.0
%2	0.0	%3	0.0	%4	0.0
%3	1.35400	%4	0.0	%5	0.0
%4	0.22100	%5	2.32700	%6	0.0
%5	0.0	%6	0.77300	%7	0.0
%6	1.87800	%7	2.39900	%8	0.0
%7	0.0	%8	0.42400	%9	0.0
%8	0.0	%9	0.0	1	0.27900
%9	0.0	1	−1.80000	2	0.0
1	−0.38550	2	0.0	3	0.0
2	1.39700	3	−0.30400	4	0.0
3	2.32100	4	−0.33650	5	0.0
4	2.24600	5	−4.49700	B	0.0
5	−1.54250	B	1.46050	C	0.0
B	0.0	C	5.26050	Al	2.34800
C	1.81750	Al	1.20750	As	0.0
Al	1.34150	Cr	−0.80100	Cd	0.0
As	1.52500	N	1.88350	Cr	0.0
Cd	0.0	O	0.06450	Ga	0.0
Cr	2.72100	Mo	6.59900	N	−0.20200
Ga	1.23950	Ni	0.0	O	−2.07300
N	1.93350	Si	−1.40950	Mo	0.0
O	0.12700	.	−0.27900	Ni	0.0
Mo	7.07400			Pt	0.0
Pt	0.0			Si	−0.38350
Si	−2.26050			.	0.52700
.	−0.35950				

$$\begin{aligned} &\mathrm{LgTC} = 0.8914\,(\pm 0.010) + 0.1380\,(\pm 0.0017) * \mathrm{DCW}\,(1, 55) \qquad (3)\\ &n = 12,\ r^2 = 0.9598,\ q^2 = 0.9490,\ s = 0.120,\ F = 239\ \text{(sub-training set)}\\ &n = 15,\ r^2 = 0.8505,\ s = 0.239\ \text{(calibration set)}\\ &n = 12,\ r^2 = 0.8811,\ s = 0.317\ \text{(test set)}\\ &n = 12,\ r^2 = 0.9295,\ s = 0.456\ \text{(validation set)} \end{aligned}$$

$$\begin{aligned} &\mathrm{LgTC} = 0.8700\,(\pm 0.0284) + 0.2283\,(\pm 0.0145) * \mathrm{DCW}\,(3, 57) \qquad (4)\\ &n = 14,\ r^2 = 0.6742,\ q^2 = 0.5570,\ s = 0.327,\ F = 49\ \text{(sub-training set)}\\ &n = 13,\ r^2 = 0.8463,\ s = 0.379\ \text{(calibration set)}\\ &n = 12,\ r^2 = 0.8508,\ s = 0.247\ \text{(test set)}\\ &n = 12,\ r^2 = 0.6751,\ s = 0.321\ \text{(validation set)} \end{aligned}$$

The calculations were carried out for three various splits into the sub-training, calibration, test, and external (invisible) validation sets. The statistical quality of these models is different, but can be estimated as satisfactory for all three splits. It is to be noted the approach is flexible, since one can modify the representation of MEMS (or the representation for other complex substances) by adding other technological attributes if necessary (Table 1). In fact the suggested approach is paradigm "response of a system is a mathematical function of group of impacts upon system".

3 Conclusions

CORAL software [16] can be used to build up predictive model of thermal conductivity of MEMS. The statistical quality of the obtained results is nearly the same for various splits of the data carried out in this study. The input data for the applied approach can be modified for other MEMS and/or other substances with complex architecture.

Acknowledgments We thank EC project PreNanoTox (contract 309666) and the EC project NANOPUZZLES (Project Reference: 309837) and EC project NanoBRIDGES (contract 295128) and the National Science Foundation (NSF/CREST HRD-0833178, and EPSCoR Award #: 362492-190200-01/NSFEPS-090378) for financial support.

References

1. G. Melagraki, A. Afantitis, H. Sarimveis, P.A. Koutentis, G. Kollias, O. Igglessi-Markopoulou, Mol. Divers. **13**, 301–311 (2009)
2. G. Melagraki, A. Afantitis, O. Igglessi-Markopoulou, A. Detsi, M. Koufaki, C. Kontogiorgis, D.J. Hadjipavlou-Litina, Eur. J. Med. Chem. **44**, 3020–3026 (2009)
3. J. García, P.R. Duchowicz, M.F. Rozas, J.A. Caram, M.V. Mirífico, F.M. Fernández, E.A. Castro, J. Mol. Graph. Model. **31**, 10–19 (2011)
4. J.C. Garro Martinez, P.R. Duchowicz, M.R. Estrada, G.N. Zamarbide, E.A. Castro, Int. J. Mol. Sci. **12**, 9354–9368 (2011)
5. E. Ibezim, P.R. Duchowicz, E.V. Ortiz, E.A. Castro, Chemometr. Intell. Lab. **110**, 81–88 (2012)
6. L.M.A. Mullen, P.R. Duchowicz, E.A. Castro, Chemometr. Intell. Lab. **107**, 269–275 (2011)
7. P.K. Ojha, I. Mitra, R.N. Das, K. Roy, Chemometr. Intell. Lab. **107**, 194–205 (2011)

8. A. Gajewicz, B. Rasulev, T.C. Dinadayalane, P. Urbaszek, T. Puzyn, D. Leszczynska, J. Leszczynski, Adv. Drug Deliv. Rev. **64**, 1663–1693 (2012)
9. A.P. Toropova, A.A. Toropov, E. Benfenati, G. Gini, Cent. Eur. J. Chem. **9**, 75–85 (2011)
10. A.A. Toropov, N.L. Voropaeva, I.N. Ruban, SSh Rashidova, Polym. Sci. A **41**, 975–985 (1999)
11. A.A. Toropov, A.P. Toropova, I. Raska Jr, E. Benfenati, G. Gini, Struct. Chem. **23**, 1891–1904 (2012)
12. J. Leszczynski, Nat. Nanotech. **5**, 633–634 (2010)
13. A.A. Toropov, D. Leszczynska, J. Leszczynski, Comput. Biol. Chem. **31**, 127–128 (2007)
14. A.A. Toropov, D. Leszczynska, J. Leszczynski, Mater. Lett. **61**, 4777–4780 (2007)
15. MEMSnet, https://www.memsnet.org/material/. Accessed 9 Feb 2013
16. CORAL, http://www.insilico.eu/coral. Accessed 19 Feb 2013
17. A.P. Toropova, A.A. Toropov, E. Benfenati, G. Gini, D. Leszczynska, J. Leszczynski, J. Comput. Chem. **32**, 2727–2733 (2011)