Journal of Molecular Graphics and Modelling 31 (2011) 57–64

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

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

A quantum mechanical quantitative structure–property relationship study of the melting point of a variety of organosilicons

Yi Liu, Andrew J. Holder*

University of Missouri-Kansas City, Department of Chemistry, Kansas City, MO 64110, United States

ARTICLE INFO

Article history:
Received 24 March 2011
Received in revised form 5 August 2011
Accepted 7 August 2011
Available online 25 August 2011

Keywords:
QSPR
Quantum mechanical
Semiempirical
Organosilicons
Melting point

ABSTRACT

We have developed quantitative structure–property relationship (QSPR) models that correlate the melting points of chain and cyclic silanes and siloxanes with their molecular structures. A comprehensive correlation was derived for a variety of molecules, but the quality of the comprehensive model was modest at best. This provided the impetus for the development of two additional models focused on silanes and siloxanes, respectively. Statistical analyses confirm the robustness of the refined models, and the chemical interpretation of the descriptors was consistent with effects expected for melting.

1. Introduction

Modern dental composite materials are a blend of glass or ceramic particles dispersed in a photo-polymerizable synthetic organic resin matrix. Currently, the most widely used resins are methylmethacrylate-based polymers that were introduced for clinical use by Bowen in the 1960s [1–3]. Biocompatibility of unreacted monomers leaching from the restorative is of continuing concern. Typical methylmethacrylate monomers, such as *bis*-GMA (2,2-*bis*[4-(2′-hydroxy-3′-methacryloxy-propoxy)phenyl]propane), HEMA (2-hydroxyehtyl methacrylate), TEGDMA (triethyleneglycol dimethacrylate), and UDMA (urethane dimethacrylate), have shown cytotoxic, mutagenic, or estrogenic effects in various studies [4–9]. In comparison, some new silorane-based resins developed by 3M-ESPE [10] exhibit lower cytotoxic and mutagenic activity than methylmethacrylate resins [11,12]. However, silorane-based polymers also suffer polymerization volume shrinkage as do virtually all commonly used dental restorative monomers. Our purpose here is to develop new silorane-based polymers that mitigate volume shrinkage, while retaining favorable biocompatibility characteristics.

Quantitative structure–property relationship (QSPR) methods have proven to be quite helpful in this regard by facilitating rational design of new compounds [13]. They not only serve as screening tools for candidate materials, but also provide suggestions on how a molecule's structure might be modified in order to enhance favorable characteristics and mitigate unfavorable characteristics. We have recently developed several QSPR models to predict a variety of chemical and physical properties and behaviors, including polymerization volume change [14], refractive index [15], lipophilicity [16], skin sensitization [17], mutagenicity [18], and flexural modulus [19].

We here introduce an extension of that previous work, a QSPR model to predict the melting points of silicon-containing materials. Melting point is closely related to a liquid's viscosity [20] and may suggest the "mixability" of a material when blended with the other components of a restorative material. Moreover, melting point affects the compound's toxicity as it is intrinsically related to solubility [21] and adequate solubility is a prerequisite for a compound to exhibit toxic effects [22]. Even more basic than these quantities though, is the simple prediction of whether a candidate monomer will be a liquid at room or physiological temperature.

Melting point is a fundamental physical property of any compound, but establishing reasonably predictive models from a structure–property relationships approach has been quite challenging. The primary difficulty is that the molecular descriptors normally used for QSPR treatments are calculated from single molecular structures (essentially gas-phase), and they may not satisfactorily describe solid state effects, such as crystal packing and condensed phase intermolecular forces [23–25]. Melting point models that account for the difference between molecules in the solid and liquid states are not easy or obvious, especially when the intrinsic complexities of the melting process such as liquid crystallinity [26] and polymorphism [27] are considered. Nevertheless,

* Corresponding author. Tel.: +1 816 235 2293; fax: +1 816 235 5502.
E-mail address: holdera@umkc.edu (A.J. Holder).

1093-3263/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2011.08.003

a number of publications have appeared [22,28–39]. The quality and predictive ability of these models vary substantially. In general, models that are derived from training sets with significant structural variation are poor ($R^2 < 0.7$) [36–39]. Not unexpectedly, more reliable models ($R^2 > 0.8$) have been derived for homologous sets such as branched vs. un-branched alkanes [33]. The best correlation ($R^2 = 0.9980$, $s = 0.51$, $n = 24$) is for *n*-alkanes [29]. Other studies utilized more sophisticated search algorithm such as artificial neural networks or genetic algorithms, and resulted in more statistically robust correlations [28,38], but at the price of coherence. We prefer traditional linear regression because the chemical significance of the descriptors in such models is easier to interpret and the information from these correlations is more suitable to guide materials discovery investigations. The molecules under examination here contain a variety of atom types including Si, C, H, O, N, and halides. They also represent a relatively diverse collection of chemical structures including chain silanes, chain siloxanes, cyclic silanes, cyclic siloxanes, and their alkyl-, aryl-, alkoxy-, aryloxy-, hydroxyl- and amino-substituted derivatives.

2. Computational methods and data treatment

The data set used in this study is comprised of silicon-containing organic compounds from various sources [40–43]. All melting points were verified against the CAS registry and NIST database [44]. Data points were further verified from original references if inconsistencies were noted. Several errors in the CAS registry were located in this process and reported to CAS. If inconsistencies remained present for a particular material, the average of the two boundary values was used if the difference was less than 10 °C. Application of these criteria resulted in a set of 129 compounds, including 25 chain siloxanes, 79 chain silanes, 20 cyclic siloxanes, and 5 cyclic silanes (Table 1). Molecular structures were optimized using the SAM1 [45] semiempirical method as implemented in *AMPAC 8.16 with Graphical User Interface* (Semichem Inc., Shawnee Mission, KS). Searches and optimization of various molecular geometries were performed to locate the lowest energy conformation of each molecule for use in the next step.

CODESSA (Semichem Inc., Shawnee Mission, KS) then calculated more than 600 molecular descriptors for each member of the set. A heuristic algorithm [46] was employed to locate the most significant descriptors and the best candidate correlations, requiring a threshold of colinearity less than 0.8. Data was exported to *SPSS* (SPSS Inc., Chicago, IL), which was used for statistical analysis of the descriptors and models. We develop QSPR models by incrementing the number of descriptors until criteria for significance for a candidate descriptor was below our tolerance (>0.01) and no more of the variation in the data can be accounted for. Each model is evaluated for pair-wise colinearity to ensure that all descriptors are statistically relevant. The best model is then selected from the candidates suggested by *CODESSA*, considering the coefficient of multiple determination (R^2), adjusted R^2 (R^2_{adj}), cross-validated R^2 (R^2_{cv}), variation inflation factor (VIF), *p* significance level and *F*-ratio in combination with the chemical sense of the descriptors as they are related to the mechanism of melting.

Outlier diagnostics were carried out following an M-estimation routine as incorporated in *SAS 9.2* (SAS Institute Inc.: Cary, NC). M-estimation was first introduced by Huber [47] in an effort to produce a more robust least squares method. In brief, M-estimation minimizes the sum of squares of the weighted residuals, where an observation with a higher residual is assigned a lower weight. The weights of each observation and the coefficients of each independent variable are obtained through an iterative process. Here, the Huber weight function was used, and an observation was treated as an outlier when its standardized robust residual was larger than 3. It must be noted that the exclusion of an outlier from the training set sometimes resulted in significant changes in the *nature* of the descriptors found through the heuristic search. Therefore, an alleged outlier in the beginning could possibly become useful after excluding the other outliers and re-training the model. In order to avoid this, only one worst outlier was excluded after each outlier diagnostic (Fig. 1), and all alleged outliers were tested in the final model to ensure they were really outside of the model's scope.

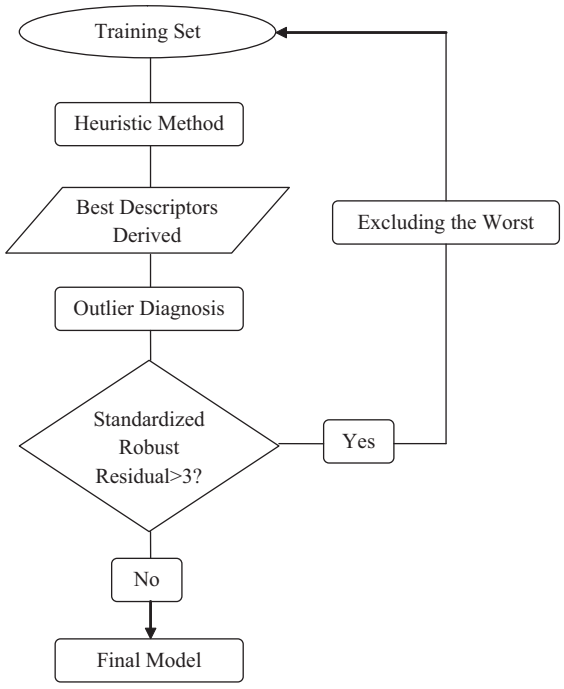


Fig. 1. Scheme of the model development including outlier diagnosis.

3. Results and discussion

3.1. Comprehensive model (Model I)

The data set of all 129 CAS/NIST-verified organosilicons was divided into a training set of 97 compounds and a test set of 32 compounds. The division was based on similar functionalities (chain/cyclic, silane/siloxane, alkyl-/alkoxy-/aryl-/aryloxy-/halide-/hydroxyl-/amino-substituted) with random assignment between the two sets. Outlier diagnosis indicated no presence of outliers in the best six-descriptor correlation model, but the model provided only modest training set statistics of $R^2 = 0.789$, $R^2_{adj} = 0.774$, $R^2_{cv} = 0.760$, and *F*-ratio = 54.2, and validation set statistics of $R^2 = 0.710$, and RMSE = 42.44 (Table 2 and Fig. 2). The six descriptors involved in this model are listed in Table 3.

3.2. Model for silanes only (Model II)

Melting point correlations for the silane subset were developed in an effort to increase the predictive power of the model for these important compounds. The 84 silanes were divided into a training set of 62 and a validation set of 19, with 3 identified as outliers (Fig. 3). Using the same procedure for descriptor screening and correlation selection, a four-descriptor correlation gave $R^2 = 0.889$, $R^2_{adj} = 0.881$, $R^2_{cv} = 0.868$, $F = 113.7$ (Table 2 and Fig. 2). The descriptors involved in Model II are listed in Table 3.

Table 1
CAS/NIST verified melting points.

CAS number	Exp. MP	Model I (comprehensive)		Model II (for silanes)		Model III (for siloxanes)	
		Calc. MP	Pred. error	Calc. MP	Pred. error	Calc. MP	Pred. error
Silanes							
1000-70-0	−23.00	7.10	30.10	0.62	23.62		
10025-78-2	−127.35	−60.03	67.32	−106.32	21.03		
10026-04-7	−66.85	−32.70	34.15	−71.02	−4.17		
1020-84-4	96.50	44.37	−52.13	121.79	25.29		
1066-40-6	−4.50	−8.26	−3.76	−16.25	−11.75		
1066-42-8	98.50	79.99	−18.51	56.01	−42.49		
107-37-9	35.00	−15.39	−50.39	−55.48	−90.48		
1111-74-6	−149.50	−143.12	6.38	−154.46	−4.96		
1112-48-7	−49.50	−64.17	−14.67	−76.10	−26.60		
112164-21-3	289.00	273.90	−15.10	313.97	24.97		
115-21-9	−106.75	−58.69	48.06	−76.73	30.02		
1174-72-7	50.50	73.30	22.80	164.73	114.23		
13435-12-6	53.00	−10.94	−63.94	78.96	25.96		
13465-78-6	−118.00	−163.28	−45.28	−126.36	−8.36		
13465-84-4	121.50	59.05	−62.45	113.25	−8.25		
13835-81-9	−25.50	37.60	63.10	6.36	31.86		
1450-14-2	11.70	−3.55	−15.25	−32.48	−44.18		
14630-40-1	23.90	−48.94	−72.84	−16.68	-40.58		
1516-80-9	63.50	45.87	−17.63	88.67	25.17		
1529-17-5	−55.00	9.61	64.61	−10.62	44.38		
1586-73-8	70.50	−18.91	−89.41	41.96	−28.54		
16881-77-9	−136.00	−80.56	55.44	−99.28	36.72		
1719-53-5	−99.78	−60.23	39.55	−86.87	12.91		
1789-58-8	−102.59	−75.28	27.31	−114.17	−11.58		
18077-14-0	−80.10	−48.25	31.85	−1.89	78.21		
18089-64-0	−109.50	−95.40	14.10	−99.16	10.34		
18171-59-0	−48.00	−24.24	23.76	−73.53	−25.53		
18395-90-9	−15.00	−35.02	−20.02	−3.37	11.63		
18666-40-5	80.50	78.08	−2.42	92.58	12.08		
18666-68-7	63.00	44.17	−18.83	70.10	7.10		
18762-95-3	127.50	88.49	−39.01	114.37	−13.14		
18844-07-0	232.50	232.65	0.15	235.96	3.46		
2117-28-4	−71.00	−47.31	23.69	−26.01	44.99		
2182-66-3	−12.50	−40.15	−27.65	−32.20	−19.70		
2345-72-4	23.75	74.61	50.86	24.80	1.05		
2441-21-6	−88.00	−76.63	11.37	−35.60	52.40		
2487-90-3	−114.90	−67.49	47.41	−85.36	29.54		
2504-64-5	27.50	79.25	51.75	35.35	7.85		
2754-32-7	136.50	114.10	−22.40	153.78	17.28		
2857-97-8	−43.35	−72.28	−28.93	−77.52	−34.17		
291-27-0	10.00	−95.62	−105.62	−30.46	−40.46		
33432-30-3	87.00	66.70	−20.30	50.70	−36.30		
3440-02-6	−23.00	37.50	60.50	58.69	81.69		
35342-88-2	81.00	−6.88	−87.88	71.03	−9.97		
353-89-9	−105.00	−89.49	15.51	−80.54	24.46		
358-06-5	−78.70	−51.76	26.94	−93.96	−15.26		
368-47-8	−18.00	−1.21	16.79	−23.59	−5.59		
373-74-0	−72.80	−137.56	−64.76	−75.47	−2.67		
379-50-0	61.50	65.08	3.58	69.85	8.35		
4095-09-4	−28.25	−38.54	−10.29	−36.05	−7.80		
4109-96-0	−122.00	−103.81	18.19	−141.85	−19.85		
4147-89-1	16.50	14.34	−2.16	39.45	22.95		
4215-80-9	53.50	59.91	6.41	83.56	30.06		
460-55-9	−122.00	−69.50	52.50	−66.27	55.73		
4648-54-8	−95.00	−31.66	63.34	−81.15	13.85		
4667-99-6	−51.00	−48.77	2.23	−55.51	−4.51		
4775-57-9	202.00	156.92	−45.08	142.93	−59.07		
5314-52-3	−99.80	−106.62	−6.82	−120.92	−21.12		
542-91-6	−134.30	−114.34	19.96	−164.54	−30.24		
562-90-3	111.00	−7.58	−118.58	41.60	−69.40		
631-36-7	−83.79	−70.08	13.71	−98.94	−15.15		
681-84-5	1.25	−57.89	−59.14	−43.19	−44.44		
75-54-7	−92.50	−92.13	0.37	−114.16	−21.66		
75-76-3	−99.00	−99.56	−0.56	−105.89	−6.89		
75-94-5	−94.90	−48.07	46.83	−59.03	35.87		
76358-47-9	51.00	10.12	−40.88	65.25	14.25		
76750-22-6	166.05	54.60	−111.45	164.90	−1.15		
778-25-6	166.00	61.92	−104.08	77.79	−88.21		
7783-29-1	−87.15	−86.24	0.91	−80.65	6.50		
7789-57-3	−73.50	−18.04	55.46	−63.72	9.78		
7789-66-4	5.15	38.97	33.82	−9.69	−14.84		
7803-62-5	−184.85	−233.77	−48.92	−155.56	29.29		
78-10-4	−78.25	−43.83	34.42	−62.35	15.90		

Table 1 (*Continued*)

CAS number	Exp. MP	Model I (comprehensive)		Model II (for silanes)		Model III (for siloxanes)	
		Calc. MP	Pred. error	Calc. MP	Pred. error	Calc. MP	Pred. error
78-62-6	−87.00	−51.09	35.91	−81.94	5.06		
791-29-7	67.30	41.48	−25.82	64.82	−2.48		
80-10-4	−22.00	50.65	72.65	35.79	57.79		
992-94-9	−155.75	−181.57	−25.82	−142.27	13.48		
993-00-0	−134.50	−131.98	2.52	−149.88	−15.38		
993-07-7	−135.90	−114.80	21.10	−125.03	10.87		
994-30-9	−75.93	−58.67	17.26	−97.85	−21.92		
998-30-1	−170.00	−54.14	115.86	-98.56	71.44		
17760-13-3	115.50	89.80	−25.70	Outliers for Model II			
18162-48-6	91.50	−47.12	−138.62	Outliers for Model II			
18171-74-9	98.50	−43.46	−141.96	Outliers for Model II			
Siloxanes							
107-46-0	−66.50	−49.59	16.91			−51.26	15.24
107-50-6	−29.00	−14.59	14.41			−36.89	−7.89
107-51-7	−83.00	−40.28	42.72			−87.70	−4.70
107-52-8	−59.00	−43.14	15.86			−76.78	−17.78
141-62-8	−76.00	−40.06	35.94			−87.81	−11.81
141-63-9	−82.00	−41.04	40.96			−85.74	−3.74
14986-21-1	−34.00	19.00	53.00			−2.07	31.93
15901-49-2	−2.84	4.50	7.34			−5.53	−2.69
17082-46-1	−155.00	−92.21	62.79			−90.95	64.05
17704-22-2	−138.00	−58.40	79.60			−86.77	51.24
17928-28-8	−80.00	−46.30	33.70			−90.84	−10.84
17940-63-5	−122.00	−10.58	111.42			−46.28	75.72
18304-82-0	−121.00	−61.02	59.98			−94.66	26.34
18754-97-7	81.25	71.43	−9.82			35.91	−45.34
18766-38-6	−53.15	−39.55	13.60			−22.90	30.25
18772-36-6	−3.15	−37.40	−34.25			−28.23	−25.08
18817-05-5	170.50	180.36	9.86			175.10	4.60
18825-68-8	170.50	93.21	−77.29			138.27	−32.23
18919-94-3	−43.15	−37.62	5.53			−15.19	27.96
2031-79-0	11.95	−4.78	−16.73			−37.34	−49.29
2370-88-9	−65.00	−95.20	−30.20			−87.26	−22.26
2401-73-2	−37.50	−34.08	3.42			−58.25	−20.75
2554-06-5	−43.50	−34.73	8.77			−63.24	−19.74
2627-95-4	−99.70	−57.84	41.86			−116.53	−16.83
31323-44-1	−47.00	48.02	95.02			−44.20	2.80
3555-47-3	−60.00	−54.32	5.68			−84.18	−24.18
4617-27-0	−22.00	−36.13	-14.13			−38.34	−16.34
540-97-6	−3.00	−13.77	−10.77			−37.21	−34.21
541-01-5	−78.00	−49.11	28.89			−69.58	8.42
541-02-6	−41.00	−7.08	33.92			−18.60	22.40
541-05-9	61.75	5.10	−56.65			40.03	−21.72
546-56-5	201.00	173.46	−27.54			197.49	−3.51
556-67-2	17.90	0.01	−17.89			−5.73	−23.63
556-68-3	30.75	−27.06	−57.81			−33.57	−64.32
556-71-8	−23.15	−25.28	−2.13			−31.13	−7.98
6166-86-5	−108.00	−122.28	−14.28			−87.23	20.77
7065-59-0	102.50	84.76	−17.74			108.41	5.91
71203-43-5	81.85	162.08	80.23			88.25	6.40
7623-01-0	−40.00	−5.68	34.32			−30.38	9.62
799-53-1	50.00	29.59	−20.41			81.27	31.27
807-28-3	45.50	57.26	11.76			115.52	70.02
1118-15-6	65.00	64.03	−0.97			−51.26	15.24
3353-68-2	46.00	132.66	86.66			−36.89	−7.89
107-46-0	−66.50	−49.59	16.91	Outlier for Model III			
107-50-6	−29.00	−14.59	14.41	Outlier for Model III			

Table 2
Model statistics for Models I, II, and III.

n	Training						External validation		
	# of descriptors	R^2	R^2_{adj}	R^2_{cv}	*F*	RMSE	*n*	R^2	RMSE
Model I									
97	6	0.742	0.725	0.705	54.2	50.45	32	0.710	42.44
Model II									
62	4	0.889	0.881	0.868	113.8	34.09	19	0.858	38.44

n	Training					LGO cross-validation	
	# of descriptors	R^2	R^2_{adj}	R^2_{cv}	*F*	RMSE	R^2	RMSE
Model III								
43	4	0.872	0.858	0.839	64.5	30.28	0.814	36.66

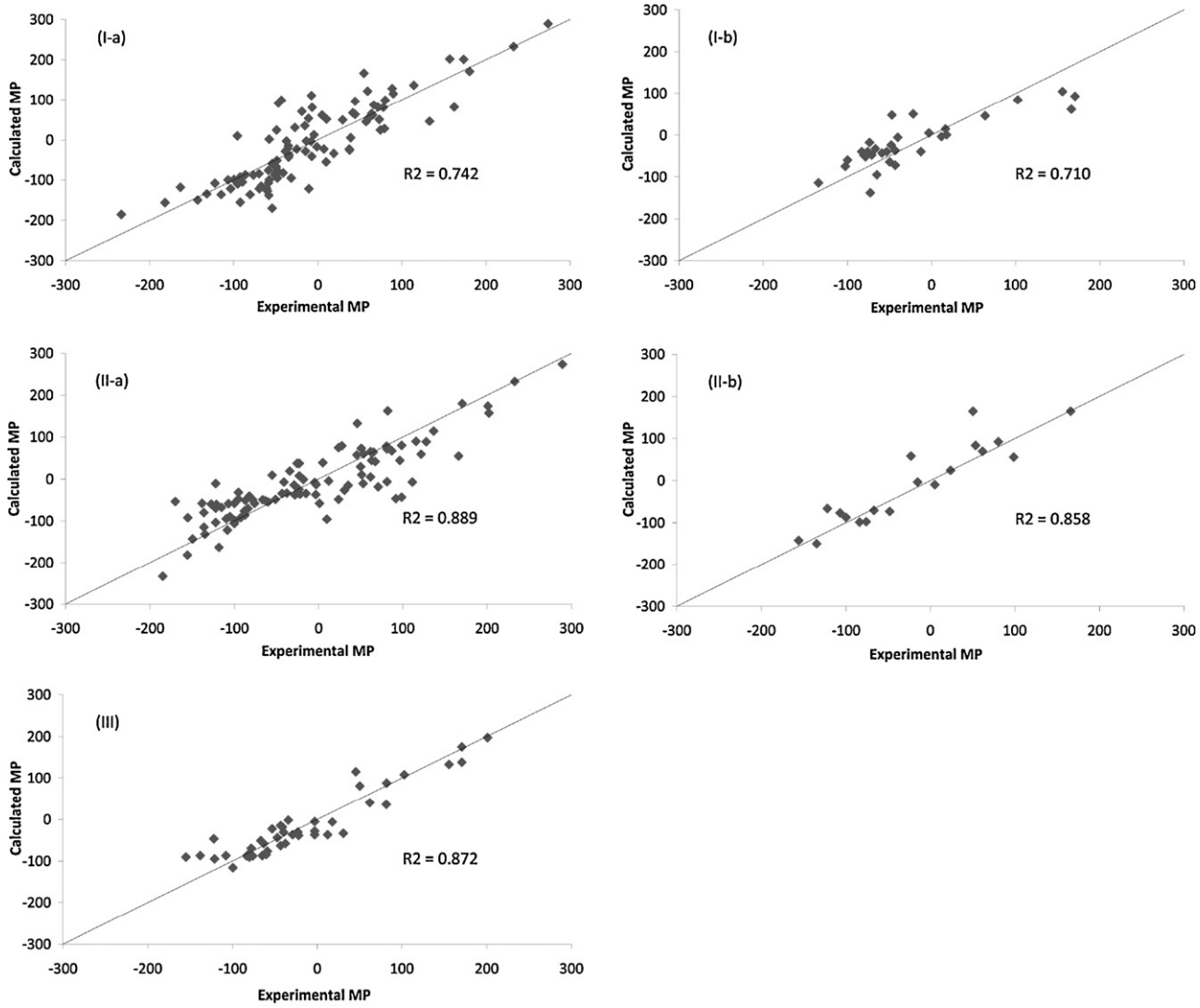
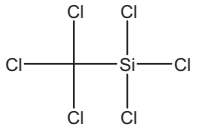


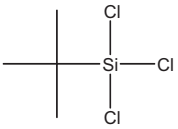
Fig. 2. Scatter diagrams of the QSPR Models (I, II and III) for (a) training and (b) validation sets.

Table 3
Descriptors in Models I, II, and III.

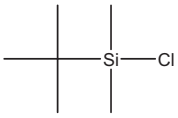
	X	ΔX	t-Test	p	β	VIF	Name of descriptors
Model I: $MP = 88.27 + 40.35d_1 - 14.89d_2 + 57.79d_3 + 86.63d_4 + 29.06d_5 + 50908.50d_6$							
	88.27	64.99	1.358	0.178			Intercept
d_1	40.35	6.025	6.696	0.000	0.648	3.273	Number of rings
d_2	−14.89	2.448	−6.082	0.000	−0.584	3.223	WNSA-3 [Zefirov's PC]
d_3	57.79	14.091	4.101	0.000	0.487	4.930	FNSA-2 [quantum-chemical PC]
d_4	86.63	13.378	6.475	0.000	0.459	1.757	Vib heat capacity (300 K)/# of atoms
d_5	29.06	6.143	4.730	0.000	0.334	1.747	HOMO energy
d_6	50908.50	12723.7	4.001	0.000	0.221	1.062	HA dependent HDCA-2 [Zefirov's PC]
Model II: $MP = -55.07 + 0.11d_1 + 2.90d_2 - 106.51d_3 - 14.27d_4$							
	−55.07	27.219	−2.023	0.048			Intercept
d_1	0.11	0.010	11.442	0.000	0.822	2.642	Gravitational index (all bonds)
d_2	2.90	0.482	6.007	0.000	0.271	1.043	ESP-HASA [quantum-chemical PC]
d_3	−106.51	26.348	−4.042	0.000	−0.257	2.064	ESP-RPCG [quantum-chemical PC]
d_4	−14.27	3.759	−3.797	0.000	−0.218	1.683	Kier flexibility index
Model III: $MP = -133.10 + 4120.60d_1 - 1872.03d_2 + 6.01d_3 + 26986.00d_4$							
	−133.10	18.02	−7.3883	0.000			Intercept
d_1	4120.60	290.88	14.166	0.000	1.353	2.702	Relative number of rings
d_2	−872.03	177.11	−4.9238	0.000	−0.398	1.941	ESP-FPSA-3 [quantum-chemical PC]
d_3	6.01	1.48	4.0642	0.000	0.289	1.498	Kier shape index (order 3)
d_4	26986.00	8665.5	3.1142	0.004	0.202	1.245	Principal moment of inertia A/# of atoms



trichloro(trichloromethyl)silane
CAS: 17760-13-3
Exp. mp: 115.5
Pred. mp: -41.53 (Model II)



t-butyltrichlorosilane
CAS: 18171-74-9
Exp. mp: 98.5
Pred. mp: -26.40 (Model II)



t-butylchlorodimethylsilane
CAS:18162-48-6
Exp. mp: 91.5
Pred. mp: -54.45 (Model II)

Fig. 3. Outliers of Model II.

The first descriptor, gravitational index (all bonds) is a constitutional descriptor that can be interpreted as primarily accounting for intermolecular dispersion forces [46]. It is defined as:

$$G_2 = \sum_{i<j}^{N_B} \frac{m_i \cdot m_j}{r_{ij}^2}$$

where m_i, m_j, r_{ij} are the atomic masses and interatomic distance (respectively), and the sum runs over all bonded atom pairs. The magnitude of the β value of this descriptor suggests that gravitational index is the dominant factor affecting the melting point of a silane. That is conceptually satisfying as heavier molecules with more polarizable electrons melt at higher temperatures in the absence of other factors.

The second descriptor is ESP (electrostatic potential) hydrogen-bond acceptor's surface area (ESP-HASA). Descriptors of this class were present in almost every boiling point and melting point model [24,34,38,48,49], unless the model was exclusively developed for molecules not capable of H-bonding. ESP-HASA is positively correlated to melting point, as expected from the influence of strong intermolecular forces such as hydrogen bonds.

The third descriptor, ESP relative positive partial charge (ESP-RPCG) based on quantum-chemical partial charges is the ratio between the partial charge of the most positive atom to the total positive charge of the molecule. It encodes information about the anisotropic distribution of positive charge in a molecule, encoding the dipole moment and point-charge effects. Modarresi's model [38] also had a similar descriptor (see below) and he attributed its contribution and negative coefficient to the presence of nodes of instability in the lattice caused by these point charges. Molecules with higher RPCG values are bound by stronger electrostatic interactions, but the interactions are also more unidirectional. As the crystal lattice starts to break down, the unidirectional electrostatic attraction diminishes rapidly with molecular movements, causing instability in the hypothetical partial melt phase, which in turn induces melting at a lower temperature. A noted above, Modarresi's [38] melting point model had a similar descriptor, the relative *negative* partial charge (RNCG), which bore the same negative sign as ESP-RPCG in our Model II.

The last descriptor is topological, the Kier flexibility index (Φ) [50]. This H-depleted molecular graph-based index is defined as:

$$\Phi = \frac{{}^1\kappa^2\kappa}{N_A}$$

In this definition, the first-order alpha-modified Kier shape index ${}^1\kappa$ encodes the number of heavy atoms in a molecule, whereas the second-order ${}^2\kappa$ encodes its degree of branching and spatial density, and N_A is the number of vertices of the graph. Overall, the Kier flexibility index is reduced when the molecule has fewer heavy atoms or a higher degree of branching. Thermodynamically speaking, a compound's melting point is determined by the ratio between its enthalpy of fusion and entropy of fusion, so more flexible molecules have a higher entropy gain as they melt. In this connection, the negative coefficient of the Kier flexibility index makes sense.

The external validation using the 19 previously excluded silanes resulted in a model where $R^2 = 0.856$ and RMSE = 37.00 (Table 2 and Fig. 2). This is a slight improvement over the comprehensive model (Model I), but has more relevant descriptors for these particular compounds.

3.3. Outliers

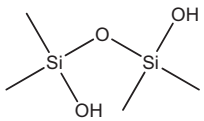
As can be seen in Fig. 3, all three outliers, trichloro-(trichloromethyl)silane (CAS: 17760-13-3), *t*-butyltrichlorosilane (CAS: 18171-74-9) and *t*-butylchlorodimethylsilane (CAS: 18162-48-6) have experimental melting points that are much higher than those predicted by the model. These three molecules are strikingly similar structurally, all of them being short-chain silanes with highly symmetric geometry and evenly distributed partial charges. This may lead to additional packing forces that are not adequately described by the molecular descriptors we used, and as a result their predicted melting points are much lower than the actual values.

3.4. Model for siloxanes only (Model III)

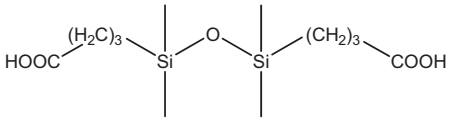
The subset of siloxanes contains 25 chain siloxanes and 20 cyclic siloxanes. Given the limited number of data points, we did not partition of the siloxane subset into training and validation sets. Instead, the leave-group-out (LGO) cross validation procedure was employed to assess the model's predictive power. The best correlation was for a four-descriptor model with $R^2 = 0.872$, $R^2_{adj} = 0.858$, $R^2_{cv} = 0.839$, $F = 64.5$ (Table 2 and Fig. 2), with 2 molecules treated as outliers. The four descriptors are provided in Table 3.

The dominant descriptor for the siloxane subset is a constitutional descriptor: the relative number of rings. A similar descriptor, the absolute number of rings, was also found significant in Modarresi's melting point model for drug-like molecules [38]. He suggested that the number of rings may represent the effect of lattice packing of a solid compound, and in agreement with Modarresi's model this descriptor bears a positive coefficient in our correlation.

The next most important descriptor is ESP fractional partial positive charged surface area-3 (FPSA3), based on quantum chemical partial charges. FPSA3 is calculated from the atomic charge weighted positive surface area (PPSA3: FPSA3 = (PPSA3/SASA) where PPSA3 $= \sum q_a^+ \cdot SA_a^+$. The negative sign of the FPSA3 regression coefficient may be related to the unique Si–O–Si fragments of siloxanes, which produce a highly positive point charge on the silicon atom($q_a{}^+$), closest to the negatively charge on oxygen. A higher FPSA3 value indicates that the oxygen atoms are less "exposed", leading to intensive Si–Si repulsion between nearby siloxane



1,1,3,3-tetramethyl-1,3-disiloxanediol
CAS:1118-15-6
Exp. mp: 65
Pred. mp: -48.47 (Model III)



1,3-bis(3-carboxypropyl)tetramethyldisiloxane
CAS:3353-68-2
Exp. mp: 46
Pred. mp: -94.51 (Model III)

Fig. 4. Outliers of Model III.

molecules. Therefore, in siloxanes with lower FPSA3 values, oxygen's charge is less shielded, favoring intermolecular electrostatic attraction, and thereby resulting in a higher melting point.

The topological index, Kier shape index (order 3) ($^3\kappa$) is related to molecular branching and flexibility. The definition of $^3\kappa$ is

$$^3\kappa \begin{cases} \dfrac{(A-3)(A-2)^2}{(^3P)^2} & \text{for even } A (A > 3) \\ \dfrac{(A-1)(A-3)^2}{(^3P)^2} & \text{for odd } A (A > 3) \end{cases}$$

where A is the number of vertices and ^3P is the number of paths with length 3, in a H-depleted molecular graph. The $^3\kappa$ values are larger when molecular branching is not present, *i.e.* when the molecule is completely linear. Less branched molecules generally demonstrate stronger dispersion forces and therefore higher melting points. This is reflected by the positive coefficient of $^3\kappa$.

The last descriptor, principal moment of inertia A/# of atoms, is related to the rotational dynamics of a molecule [51]. The principal moment of inertia A is the smallest moment of inertia when the axes are oriented in such a way that the mass distribution of the molecule is symmetrical about the axes. Considering a molecule as a rigid rotor, a high principal moment of inertia indicates that the molecule exhibits less rotational motion when thermodynamically agitated. As a result, the crystal lattice is more resistant to perturbation, and the melting point must increase to overcome this enhanced stability.

Both Models II and III have one topological index: the Kier flexibility index in Model II and the Kier shape index in Model III. These two descriptors are both derived based on molecular graph theory, and can both account for such structural features as degree of branching, geometrical symmetry, degree of saturation, and number of heavy atoms. Similar as they are, they bear coefficients with opposite signs! In Model II the Kier flexibility index has a negative coefficient, whereas in Model III the Kier shape index has a positive coefficient. This observation actually matches the widely accepted notion that, for a substance to melt, it must overcome inter-molecular forces as well as solid phase packing. Less branched and less symmetric molecules exhibit stronger dispersion or dipole–dipole forces which lead to higher melting and boiling points. On the other hand, the decreased degree of symmetry might lead to less regular solid phase packing resulting in a lower melting point. Apparently, in the case of siloxanes (Model III), the effect of the dispersion or dipole–dipole interactions outweighs that of the solid-phase packing, hence the positive sign. When compared to silane molecules, siloxane molecules have many more oxygen atoms which contribute to inter-molecular interactions.

Monte-Carlo cross-validation (MCCV) [52], a variation from the exhaustive leave-group-out cross-validation (LGO-CV) procedure, was also used to evaluate the predictive power of Model III. In the MCCV process, the total n observations are randomly partitioned into a training subset of n_c observations and a validation subset of n_v observations. The training subset (with n_c data points) is used to train the model, while the validation subset (with n_v data points, $n_v = n - n_c$) is used to test how well the model predicts the new data. The process is iterated N times. The MCCV was carried out in *Matlab 7* (The MathWorks, Natick, MA), using a practical compromise of $n_c = n_v = n/2$ and number of iteration $N = 100n$ [53,54]. The R^2 and RMSE of cross-validation are listed in Table 2. As with Model II, this is a slight improvement over the comprehensive model.

3.5. Outliers

We treated two siloxanes, 1,1,3,3-tetramethyl-1,3-disiloxanediol (CAS: 1118-15-6) and 1,3-bis(3-carboxypropyl)-tetramethyldisiloxane (CAS: 3353-68-2) (Fig. 4), as outliers. These two molecules can form hydrogen bonds, which the other members of the set cannot. When included in the training set, hydrogen-bonding associated descriptors become important. However, the values of the descriptors are actually zero for the other 43 siloxanes. This was not an issue for the comprehensive model (Model I) and silane-only model (Model II), as many hydrogen-bond-capable compounds are in the silane subset. For the separate siloxane subset, however, this produced an unstable model and poor predictive power, as evidenced by errors in cross-validation. When we predicted the two outliers' melting points by Model III, the calculated values were much lower than the experimental values (Fig. 4), as would be expected from a model that ignores strong intermolecular hydrogen-bonding effects.

4. Conclusions

In this work, QSPR models for melting points of silanes and siloxanes are presented. The descriptors involve constitutional, electrostatic, quantum chemical and thermodynamic categories, primarily encoding information about intermolecular forces and solid phase packing. This result is consistent with expectations from chemical intuition. Refined models developed specifically for silanes or siloxanes demonstrate much better quality than a comprehensive model describing both sets of molecules.

References

[1] R.L. Bowen, Dental filling material comprising vinyl silane treated fused silica and a binder consisting of the reaction product of Bisphenol and glicidyl acrylate, US Patent 3066112.
[2] R.L. Bowen, Adhesive bonding of various materials to hard tooth tissues. II. Bonding to dentin promoted by a surface-active comonomer, J. Dent. Res. 44 (1965) 895–902.
[3] R.L. Bowen, Silica-resin direct filling material and method of preparation, US Patent 3194783, 3194784.
[4] W. Geurtsen, F. Lehmann, W. Spahl, G. Leyhausen, Cytotoxicity of 35 dental resin composite monomers/additives in permanent 3T3 and three human primary fibroblast cultures, J. Biomed. Mater. Res. 41 (1998) 474–480.
[5] K.J. Soderholm, A. Mariotti, BISGMA-based resins in dentistry: are they safe? J. Am. Dent. Assoc. 130 (1999) 201–209.
[6] B. Thonemann, G. Schmalz, H.A. Hiller, H. Schweikl, Response of L929 mouse fibroblasts, primary and immortalized bovine dental papilla-derived cell lines to dental resin components, Dent. Mater. 18 (2002) 318–323.

[7] Y. Issa, D.C. Watts, P.A. Brunton, C.M. Waters, A.J. Duxbury, Resin composite monomers alter MTT and LDH activity of human gingival fibroblasts in vitro, Dent. Mater. 20 (2004) 12–20.
[8] M. Lefeuvre, K. Bourd, M.-A. Loriot, M. Goldberg, P. Beaune, A. Périanin, et al., TEGDMA modulates glutathione transferase P1 activity in gingival fibroblasts, J. Dent. Res. 83 (2004) 914–919.
[9] D.H. Lee, B.-S. Lim, Y.-K. Lee, S.-J. Ahn, H.-C. Yang, Involvement of oxidative stress in mutagenicity and apoptosis caused by dental resin monomers in cell cultures, Dent. Mater. 22 (2006) 1086–1092.
[10] R. Guggenberger, W. Weinmann, Exploring beyond meth-acrylates, Am. J. Dent. 13 (2000) 82–84.
[11] H. Schweikl, G. Schmalz, W. Weinmann, Mutagenic activity of structurally related oxiranes and siloranes in Salmonella Typhimurium, Mutat. Res. 521 (2002) 19–27.
[12] H. Schweikl, G. Schmalz, W. Weinmann, The induction of gene mutation and micronuclei by oxiranes and siloranes in mammalian cells in vitro, J. Dent. Res. 83 (2004) 17–21.
[13] A.J. Holder, K.V. Kilway, Rational design of dental materials using computational chemistry, Dent. Mater. 21 (2005) 47–55.
[14] M.D. Miller, A.J. Holder, K.V. Kilway, G.J. Giese, J.E. Finley, D.M. Travis, et al., Quantum-mechanical QSPR models for polymerization volume change of epoxides and methacrylates based on mercury dilatometry results, Polymer 47 (2006) 8595–8603.
[15] A.J. Holder, L. Ye, J.D. Eick, C.C. Chappelow, An application of QM-QSAR to predict and rationalize the refractive index of a wide variety of simple organic/organosilicon molecules, QSAR Comb. Sci. 25 (2006) 342–349.
[16] A.J. Holder, L. Ye, D.M. Yourtee, A. Agarwala, J.D. Eick, C.C. Chappelow, An application of the QM-QSAR method to predict and rationalize lipophilicity of simple monomers, Dent. Mater. 21 (2005) 591–598.
[17] M.D. Miller, D.M. Yourtee, A.G. Glaros, C.C. Chappelow, J.D. Eick, A.J. Holder, Quantum mechanical structure–activity relationship analyses for skin sensitization, J. Chem. Inf. Model. 45 (2005) 924–929.
[18] A.J. Holder, L. Ye, Quantum mechanical quantitative structure–activity relationships to avoid mutagenicity, Dent. Mater. 25 (2009) 20–25.
[19] A.J. Holder, Y. Liu, A quantum mechanical quantitative structure–activity relationship study of the flexural modulus of C, H, O, N-containing polymers, Dent. Mater. 26 (2010) 840–847.
[20] J. Nimko, J. Kukkonen, K. Riikonen, A model for evaluating physicochemical sunstance properties required by consequence analysis models, J. Hazard. Mater. 91 (2002) 43–61.
[21] N. Jain, S.H. Yalkowsky, Estimation of the aqueous solubility I: application to organic nonelectrolytes, J. Pharm. Sci. 90 (2001) 234–252.
[22] J.C. Dearden, The QSAR prediction of melting point, a property of environmental relevance, Sci. Total Environ. 109/110 (1991) 59–68.
[23] A.R. Katritzky, V.S. Lobanov, M. Karelson, QSPR. The correlation and quantitative prediction of chemical and physical properties from structure, Chem. Soc. Rev. 24 (1995) 279–287.
[24] M. Karelson, U. Maran, Y. Wang, A.R. Katritzky, QSPR QSAR models derived using large molecular descriptor spaces. A review of CODESSA applications, Collect. Czech. Chem. Commun. 64 (1999) 1551–1571.
[25] A.R. Katritzky, R. Jain, A. Lomaka, R. Petrukhin, U. Maran, M. Karelson, Perspective on the relationship between melting points and chemical structure, Cryst. Growth Des. 1 (2001) 261–265.
[26] R. Steinstrasser, L. Pohl, Chemistry and applications of liquid crystals, Angew. Chem. Int. Ed. Engl. 12 (1973) 617–630.
[27] A. Gavezzotti, M. Simonetta, Crystal chemistry in organic solids, Chem. Rev. 82 (1982) 1–13.
[28] A. Habibi-Yangjeh, E. Pourbasheer, M. Danandeh-Jenagharad, Prediction of melting point for drug-like compounds using Principal Component–Genetic Algorithm–Artificial Neural Network, Bull. Korean Chem. Soc. 29 (2008) 833–841.
[29] M.P. Hanson, D.H. Rouvary, in: R.B. King, D.H. Rouvary (Eds.), Graph Theory and Topology in Chemistry, vol. 51, Elsevier, Amsterdam, 1987, p. 201.
[30] R. Abramowitz, S.H. Yalkowsky, Melting point, boiling point and symmetry, Pharm. Res. 7 (1990) 942–947.
[31] A.R. Katritzky, E.V. Gordeeva, Traditional topological indexes vs. electronic, geometrical, and combined molecular descriptors in QSAR/QSPR research, J. Chem. Inf. Comput. Sci. 33 (1993) 835–857.
[32] R. Murugan, M.P. Grendze, J.E.J. Toomey, A.R. Katritzky, M. Karelson, V.S. Lobanov, et al., Predicting physical properties from molecular structure, CHEMTECH 24 (1994) 17–19.
[33] M. Charton, B. Charton, Quantitative description of structural effects on melting points of substituted alkanes, J. Phys. Org. Chem. 7 (1994) 196–206.
[34] A.R. Katritzky, U. Maran, M. Karelson, V.S. Lobanov, Prediction of melting points for the substituted benzenes: a QSPR approach, J. Chem. Inf. Comput. Sci. 37 (1997) 913–919.
[35] J.D. Dyekjar, S.O. Jonsdottir, QSPR models based on molecular mechanics and quantum chemical calculations 2. Thermodynamic properties of alkanes, alcohols, polyols, and ethers, Ind. Eng. Chem. Res. 42 (2003) 4241–4259.
[36] C.A.S. Bergstrom, U. Norinder, K. Luthman, P. Artursson, Molecular descriptors influencing melting point and their role in classification of solid drugs, J. Chem. Inf. Comput. Sci. 43 (2003) 1177–1185.
[37] M. Karthikeyan, R.C. Glen, A. Bender, General melting point prediction based on a diverse compound data set and artificial neural networks, J. Chem. Inf. Model. 45 (2005) 581–590.
[38] H. Modarresi, J.C. Dearden, H. Modarress, QSPR correlation of melting point for drug compounds based on different sources of molecular descriptors, J. Chem. Inf. Model. 46 (2006) 930–936.
[39] D. Zhou, Y. Alelyunas, R. Liu, Scores of entended connectivity fingerprint as descriptors in QSPR study of melting point and aqueous solubility, J. Chem. Inf. Model. 48 (2008) 981–987.
[40] V. Bazant, V. Chvalovsky, J. Rathousky, Organosilicon Compounds, Pub. House of the Czechoslovak Academy of Sciences, Prague, 1965.
[41] E.G. Rochow, Silicon and Silicones: about stone-age tools, antique pottery, modern ceramics, computers, space materials, and how they all got that way, Springer-Verlag, Berlin, New York, 1987.
[42] A.L. Smith, The Analytical Chemistry of Silicones, New York, Wiley, 1991.
[43] R.E. Kirk, D.F. Othmer, Silicon Compounds Kirk-Othmer Encyclopedia of Chemical Technology, vol. 22, J. Wiley, Hoboken, New Jersey, 2004.
[44] NIST, Chemistry WebBook.
[45] M.J.S. Dewar, C. Jie, G. Yu, SAM1: the first of a new series of general purpose quantum mechanical molecular models, Tetrahedron 49 (1993) 5003–5038.
[46] A.R. Katritzky, L. Mu, V.S. Lobanov, M. Karelson, Correlation of boiling points with molecular structure 1. A training set of 298 diverse organics and a test set of 9 simple inorganics, J. Phys. Chem. 100 (1996) 10400–10407.
[47] P.J. Huber, Robust Statistics, Wiley, 1981.
[48] A.R. Katritzky, R. Murugan, M.P. Grendze, J.E. Toomey, M. Karelson, V.S. Lobanov, et al., Predicting physical properties from molecular structure, CHEMTECH 24 (1994) 17–23.
[49] A.R. Katritzky, V.S. Lobanov, M. Karelson, R. Murugan, M.P. Grendze, J.E.J. Toomey, Comprehensive descriptors for structural and statistical analysis. 1. Correlations between structure and physical properties of substituted pyridines, Rev. Roum. Chim. 41 (1996) 851.
[50] L.B. Kier, Computational Chemical Graph Theory, Nova Science Publishers, New York, 1990.
[51] R. Todeschini, V. Consonni, Handbook of molecular descriptors, in: R. Mannhold, H. Kubinyi, H. Timmerman (Eds.), Methods and Principles in Medicinal Chemistry, vol. 11, Wiley-VCH, Weinheim, 2000, p. 1.
[52] J. Shao, Linear model selection by cross-validation, J. Am. Stat. Assoc. 88 (1993) 486–494.
[53] A.A. Toropov, B.F. Rasulev, J. Leszczynski, QSAR modeling of acute toxicity for introbenzene derivatives toward rats: comparative analysis by MLRA and optimal descriptors, QSAR Comb. Sci. 26 (2007) 686–693.
[54] D.A. Kopovalov, N. Sim, E. Deconinck, Y. Vander Heyden, D. Coomans, Statistical confidence for variable selection in QSAR models via Monte Carlo cross-validation, J. Chem. Inf. Model. 48 (2008) 370–383.