

Quantitative Property-Property Relationships (QPPRs) and Molecular-Similarity Methods for Estimating Flash Points of Si-Organic and Ge-Organic Compounds

Axel Drefahl

Owens Technology, Inc.
5355 Capital Court, Suite 106
Reno, NV 89502, U.S.A.
axel@owenstechnology.com

Abstract: Good correlations between normal boiling points and flash points of Si-organic and Ge-organic compounds have been found. These correlations are discussed and compared with known boiling-point/flash-point correlations for organic compounds. Since boiling point data are often not available, application of molecular-similarity methods for the estimation of flash points of query compounds from flash points of structurally related source compounds in a database have been explored. Relationships, called quantitative source-target differences (QSTDs), that allow the estimation of a query (new target) from database compounds (source compounds) have been developed. For example, the QSTD relationship,

$$T_f\{\text{R}_4\text{Ge}\}/^\circ\text{C} = 19.0474 + 0.9149 \cdot T_f\{\text{R}_4\text{Si}\}/^\circ\text{C} \quad (m = 13, r = 0.9832),$$

allows the estimation of the flash point (T_f) of a Ge-organic compound from the known flash point of the analogous compound that contains a Si atom instead of a Ge atom, but otherwise has the same molecular structure as the query. Further, QSTDs to estimate flash points of Si-organic compounds from related Si-organic compounds are presented. The QSTD approach additionally allows to predict lower and upper boundary values between which the true flash point of a query compound will be found with high probability. Applications of this approach with respect to flammability classification and fire hazard assessment are discussed.

Keywords: fire hazard assessment; flash point estimation; molecular similarity; silanes; germanes.

1 INTRODUCTION

Methods for estimating properties of chemical compounds include quantitative property-property relationships (QPPRs), quantitative structure-property relationships (QSPRs), and molecular-similarity methods. Their application to the estimation of various physicochemical properties for structurally diverse sets of compounds have been reviewed by Reinhard and Drefahl [1999]. QPPRs developed for estimation of flash points of organic compounds typically relate the flash point, T_f , to the normal boiling point, T_{nb} . In addition to T_{nb} , liquid density or enthalpy of vaporization have been included in multi-parameter QPPRs. Such relationships have recently been reviewed by Catoire and Naudet [2004]. Hsieh [1997] developed two QPPRs, one for silicones and one for silicones com-

bined with various organic compounds. The present study reports two-parameter QPPRs for substituted germanes and silanes. Germanes have not been included in previous QPPRs for estimation of T_f . In addition to T_{nb} , we include, as the second parameter, the total number of carbon atoms, N_C , in the molecule. This parameter was reported by Catoire and Naudet [2004] to improve correlations for organic compounds.

Observed boiling points are often missing or only reduced boiling points are available. For those compounds, T_f may be estimated using QSPRs, in which all independent variables are derived from molecular structure. Katrizky et al. [2001] and Muruga et al. [June 1994] demonstrated QSPR development with regression analysis and Tetteh et al. [1999] with a neural network approach. Applicability of both QPPRs and QSPRs depend on com-

position and structure of a query compound, which may deviate from the compositional and structural confinements of the compound set used in the derivation of a particular relationship. Relating a query directly to structurally similar compounds in a database, therefore, provides an approach that allow property estimation without ambiguity of method selection. Herein, we derive molecular-similarity methods based on the approach of molecular difference recognition by Drefahl and Reinhard [1993]. We adapt that approach to the estimation of T_f for (a) substituted germanes from T_f values of analogous silanes and for (b) substituted silanes from similar silanes differing in exactly one substituent. Substituted silanes and germanes are important precursors in the fabrication of semiconductor devices, functionalized surfaces and nanostructures. Design of novel structures, devices and synthetic routes thereto requires novel precursors, for which it is desirable to know physicochemical and environmental properties and safety parameters prior to their synthesis. Confronted with such virtual libraries of silanes and germanes, the proposed methods allow flash point estimation and flammability classification of virtual compounds based on data of related existing compounds.

2 DATA SET

The data set consists of closed-cup T_f and T_{nb} values of 13 substituted germanes and 123 substituted silanes [Drefahl, 2006]. Those compounds, for which only reduced boiling points are reported, are only included in the development of molecular-similarity methods. The substituents of the germanes and silanes include Cl, Br, alkyl, alkenyl, phenyl, alkoxy, acetyl, and acetoxy groups, where alkyl, alkenyl, and phenyl substituents may themselves contain halogen, alkyl, or phenyl substituents. The set of germanes is bounded by: $-19 \leq T_f/^\circ\text{C} \leq 160$; $43 \leq T_{nb}/^\circ\text{C} \leq 274$ and $1 \leq N_C \leq 16$. The set of silanes is bounded by: $-27 \leq T_f/^\circ\text{C} \leq 175$; $36 \leq T_{nb}/^\circ\text{C} \leq 304$ and $1 \leq N_C \leq 16$. In the following, T_f and T_{nb} always refer to temperature values in $^\circ\text{C}$. Both compound sets represent liquids ranging from very high to low fire risk.

3 METHODS

QPPRs. The following two-parameter equation is proposed, in which the coefficients a_0 , a_1 , and a_2 are derived by regression analysis separately for the

germanes and silanes:

$$T_f = a_0 + a_1 T_{nb} + a_2 N_C \quad (1)$$

Performance of this equation is compared with the equation by Catoire and Naudet [2004],

$$T_f^* = 0.3544 \cdot (T_{nb}^*)^{1.14711} \cdot N_C^{-0.07677}, \quad (2)$$

with $T_f^* = T_f + 273.15$ and $T_{nb}^* = T_{nb} + 273.15$. For silanes, (1) is also compared with the equation by Hsieh [1997],

$$T_f = -51.2385 + 0.4994 T_{nb} + 0.00047 T_{nb}^2, \quad (3)$$

that was derived from silicones including silanes (Si_1 compounds) and Si_2 to Si_8 compounds.

Molecular Similarity. Pairs of molecules, $(S_1, T_1), \dots, (S_m, T_m)$, are considered, where S_i and T_i with $1 \leq i \leq m$ are structurally similar. Any source molecule S_i is related to its target molecule T_i by the same formal transformation $S_i \rightarrow T_i$. Such a transformation, hence called source-target transformation (STT), consists in the formal replacement of an atom or a locally confined substructure in S_i by a different atom or substructure to obtain T_i . In the pair (trimethylsilane, trimethylgermane), for example, Si is formally replaced by Ge; in the pair (trimethylsilane, tetramethylsilane) a H atom is replaced by a methyl group. In another example, Figure 1 shows four pairs of silanes in which a Si-adjacent methyl group is replaced by a benzyl group resulting in an increase of T_f by between 89 and 102 $^\circ\text{C}$ (Q3 in Table 2).

We define the empirical lower STT boundary,

$$B_l^\Delta = \min\{(T_f\{T_i\} - T_f\{S_i\}) \mid 1 \leq i \leq m\}, \quad (4)$$

and the empirical upper STT boundary,

$$B_u^\Delta = \max\{(T_f\{T_i\} - T_f\{S_i\}) \mid 1 \leq i \leq m\}. \quad (5)$$

The superscript Δ indicates the association of the boundary with a molecular-structure difference described by a particular STT. For the pairs in Figure 1 we obtain $B_l^\Delta = 89^\circ\text{C}$ and $B_u^\Delta = 102^\circ\text{C}$. Further, we study relationships between source and target flash points, $T_f\{T_i\}$ and $T_f\{S_i\}$, with equation

$$T_f\{T_i\} = c_0 + c_1 T_f\{S_i\}, \quad (6)$$

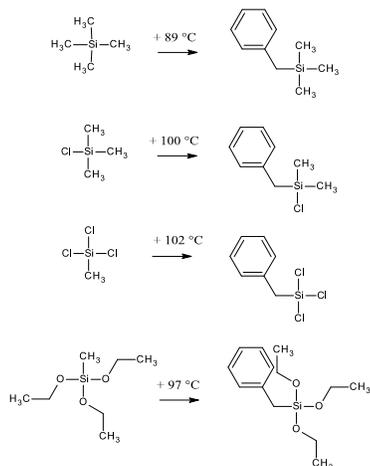


Figure 1: The formal transformation of tetramethylsilane ($T_f = -27^\circ\text{C}$), trimethylchlorosilane ($T_f = -27^\circ\text{C}$), methyltrichlorosilane ($T_f = -15^\circ\text{C}$) and methyltriethoxysilane ($T_f = 30^\circ\text{C}$) to benzyltrimethylsilane ($T_f = 62^\circ\text{C}$), benzyltrimethylchlorosilane ($T_f = 73^\circ\text{C}$), benzyltrimethyltrichlorosilane ($T_f = 87^\circ\text{C}$), and benzyltriethoxysilane ($T_f = 127^\circ\text{C}$), respectively. The associated change of T_f is indicated above the arrow.

in which c_0 and c_1 are coefficients to be derived by regression analysis for sets of compound pairs with the same STT. Relationship (6) forms a quantitative source-target difference (QSTD) method for a given STT. (6) can then be applied to estimate $T_f\{Q\}$ of a query, $Q = T_k$ with $k > m$ from source compound S_k if $S_k \rightarrow Q$ complies with that STT. Definitions (4) and (5) form a QSTD interval (QSTDI) for a given STT. Such a QSTDI, (B_l^Δ, B_u^Δ) , leads to the relation

$$T_f\{S_k\} + B_l^\Delta \leq T_f\{Q\} \leq T_f\{S_k\} + B_u^\Delta, \quad (7)$$

that provides an interval estimate of $T_f\{Q\}$, which is especially desirable when the correlation in (6) was found to be statistically unsatisfying. We derive QSTDs for germanes based on $\text{R}_4\text{Si} \rightarrow \text{R}_4\text{Ge}$ and for silanes based on transformations listed in Table 2. The STTs include substitution of a Si-adjacent H atom by a Cl atom (Q1), replacement of methyl by aryl or cyclohexyl substituents (Q2-Q6), replacement of alkyl by alkenyl substituents (Q7, Q8), and substitution of an alkyl-H atom by halogen (Q9, Q10). The substituents are denoted as A = alkyl, Me = methyl, Et = ethyl, Pr = propyl, Vyl = vinyl, Ayl = allyl, cHx = cyclohexyl, Ph = phenyl, Bz = benzyl, Phn = phenetyl, and

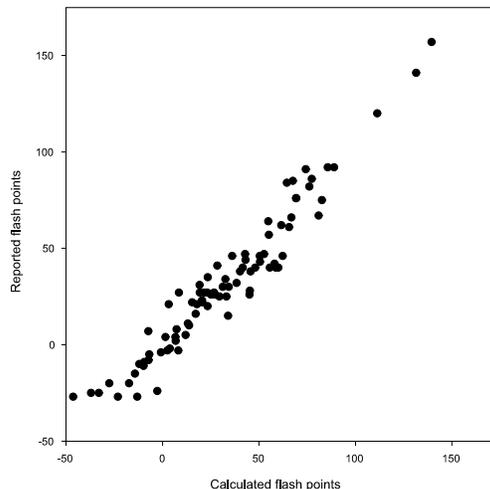


Figure 2: Comparison between calculated and reported T_f values for 86 silanes using (1) with $a_0 = -59.8301$, $a_1 = 0.7065$, and $a_2 = -1.3185$.

pTo = p - tolyl.

4 RESULTS

QPPRs. For germanes, (1) is derived by linear regression with $a_0 = -42.9177$, $a_1 = 0.5711$, and $a_2 = -1.3783$ for $n = 12$ compounds with multiple correlation coefficient $r = 0.9528$ and F -value of 44.328. This correlation coefficient shows an improvement over $r = 0.9447$ obtained for the one-parameter correlation between T_f and T_{nb} . For silanes, (1) is derived for $n = 86$ compounds with $a_0 = -59.8301$, $a_1 = 0.7065$, $a_2 = -1.3185$, $r = 0.9647$ ($r = 0.9623$ for one-parameter correlation), and $F = 556.58$. Figure 2 shows the agreement between calculated and reported T_f values. In Table 1 the mean absolute deviation (MAD) and the maximum absolute deviation (MXAD) are compared. MAD and MXAD are derived from the reported T_f values and those calculated with equations (1) through (3), as specified in the second column.

QSTDs and QSTDI. T_f values of germanes (R_4Ge) exceed T_f values of their analogues silanes (R_4Si) by $B_l^\Delta = 3^\circ\text{C}$ to $B_u^\Delta = 31^\circ\text{C}$. A QSTD correlation based on (6),

$$T_f\{\text{R}_4\text{Ge}\} = 19.0474 + 0.9149 \cdot T_f\{\text{R}_4\text{Si}\}, \quad (8)$$

Table 1: Comparison of QPPR performance for R_4X compounds (MAD and MXAD are in $^{\circ}C$).

X	Equation	n	MAD	MXAD
Ge	(1), this work	12	7.7	14.4
Ge	(2) by Catoire	12	21.0	43.1
Si	(1), this work	86	7.6	21.3
Si	(3) by Hsieh	86	7.8	29.1
Si	(2) by Catoire	86	10.6	53.0

with $m = 13$ and $r = 0.9832$. QSTDI and QSTD results for silanes are shown in Table 2 and 3, respectively. Values of the coefficients for (6) are given in parentheses when a poor correlation ($r < 0.8$) is found.

Table 2: Selected STTs with B_l^{Δ} and B_u^{Δ} values for ten different sets of silane pairs. The number of pairs, m , for each STT is given in Table 3.

Q#	Source \rightarrow Target	B_l^{Δ}	B_u^{Δ}
1	$R_3Si-H \rightarrow R_3Si-Cl$	7	52
2	$R_3Si-Me \rightarrow R_3Si-Ph$	45	106
3	$R_3Si-Me \rightarrow R_3Si-Bz$	89	102
4	$R_3Si-Me \rightarrow R_3Si-Phn$	90	106
5	$R_3Si-Me \rightarrow R_3Si-pTo$	90	107
6	$R_3Si-Me \rightarrow R_3Si-cHX$	74	106
7	$R_3Si-Et \rightarrow R_3Si-Vyl$	-7	4
8	$R_3Si-Pr \rightarrow R_3Si-Ayl$	-10	12
9	$R_3SiA-H \rightarrow R_3SiA-Cl$	-12	59
10	$R_3SiA-H \rightarrow R_3SiA-Br$	13	68

5 DISCUSSION

QPPRs. Previous studies have demonstrated the significance of T_{nb} for estimating flash points of organic compounds including silicones. Our correlation results for germanes and silanes confirm the significance of T_{nb} as a predictive descriptor. MAD and MXAD values in Table 1 show that (2) is not suitable for estimating T_f of germanes, whereas (1) especially derived with and for germanes in this work shows excellent statistical parameters. Similarly, (2) performs poorly in estimating T_f of silanes, whereas both the equation of this work and of Hsieh show a good fit between calculated and reported values. The comparison shows that QPPRs designed

Table 3: Coefficients of (6) and statistical parameters for the STTs in Table 2.

Q#	c_0	c_1	m	r
1	27.87	0.980	19	0.9660
2	77.37	0.923	23	0.9078
3	97.27	1.027	4	0.9799
4	99.07	1.052	4	0.9156
5	(106.5)	(1.786)	3	0.7815
6	(73.98)	(0.165)	4	0.1077
7	-0.953	0.949	6	0.9677
8	1.755	0.897	4	0.8793
9	(32.04)	(0.693)	17	0.6749
10	43.68	0.791	9	0.8066

with and for a particular compound class estimate T_f more accurately than general methods.

QSTDs and QSTDI. The number of germanes with reported T_f values is much smaller than the number of silanes, but missing values for many germanes can now be filled in from values of analogous silanes by using (8). ($B_u^{\Delta} - B_l^{\Delta}$) is positive for all 13 silane/germane pairs of this study suggesting that the T_f of a germane can be predicted with good confidence to be greater than the value of its silane analogue. Continuing on with replacing a IV group atom by its adjacent lower IV group atom, one may want to proceed by estimating T_f values of silanes from the organic compound analogue, a substituted methane. Although the the number of reported T_f values for organic compounds is quite large most of those silane analogues are not part of it. Since the number of available T_f values for silanes is of respectable size, however, the chance of finding a silane that differ from a query by just one substituent is good. Hence, T_f estimates for new silanes are proposed to be derived from related silanes rather than from analogous "methanes". The STTs in Table 2 and 3 were selected to demonstrated this approach. For the STTs Q1-Q4 and Q7 in Table 3 excellent correlations are found, of which the correlation for Q1 and Q2 with 19 and 23 silane pairs, respectively, are statistically significant. Some of the remaining STTs show no or poor correlations, which are not suitable for T_f estimation. For example, T_f values for methyl/cyclohexyl analogues (Q6) are uncorrelated, but T_f value of the cyclohexyl analogue is predicted to be at least by $B_u^{\Delta} = 74^{\circ}C$ higher than the T_f of the methyl-containing source. Replacement of a Si-adjacent H atom by a Cl atom (Q1) always increases T_f , whereas replacement of an alkyl-group H atom by Cl (Q9) not always does. Transformation of an alkyl into the corresponding alkenyl group (Q7

and Q8) leaves T_f approximately unchanged. Replacing a methyl group by a phenyl group (Q2) increases T_f by at least 45°C and by a phenylalkyl or alkylphenyl group (Q3-Q5) by at least 89°C. With the exception of Q7 to Q9 the selected STTs are capable of reliably identifying silanes with T_f values that are higher than the values of their source analogue. This demonstrates that QSTDs allow the identification of new compounds with flammability concern lower than that for similar database compounds, or, by switching source and target, of higher flammability concern. QSTDs are excellent methods to rank new compounds within the property values of database compounds.

Fire hazard classification. T_f data are essential in identifying and assessing fire hazards. Different systems exist to rank or classify chemicals by fire hazard. Paralikas and Lygeros [2005] and Spencer and Colonna [2002] underline that there is no single property to describe or appraise flammability and fire risk of materials. Nevertheless, some classification systems for flammability (for example, see http://www.knowledgebydesign.com/tlmc/tlmc_safety.html) rely primarily on flash point information. A typical classification system uses boundary-defined intervals to identify hazard categories for chemicals, ranking the severity of danger with numbers from zero to four: 0 if compound will not burn, 1 if $T_f > 93.4^\circ\text{C}$, 2 if $37.8^\circ\text{C} < T_f < 93.4^\circ\text{C}$, 3 if $22.8^\circ\text{C} < T_f < 37.8^\circ\text{C}$, and 4 if $T_f < 22.8^\circ\text{C}$. Relation (7) enables compound classification with respect to such systems. If B_f^Δ and B_u^Δ fall into the same interval, a single category results. Otherwise two or more neighboring categories will be identified where a fuzzy-set-like assignment for each category can be derived from the degree of coverage. For example, T_f of benzylallyloxy-dimethylsilane is estimated from $T_f = 0^\circ\text{C}$ of allyloxytrimethylsilane by using relation 7 with boundaries for transformation Q3 in Table 2 as between 89 and 102 °C. Then, a possible probability assignment predicting that benzyl-allyloxy-dimethylsilane belongs to classes 0, 1, 2, 3, and 4 could be 0, 0, 0, $(93.4 - 89)/(93.4 - 37.88) \approx 0.1$, and 0.9, respectively. Other assignment procedures and additional estimation methods for properties including the boiling point or auto-ignition temperature may be applied in estimating fire hazard. The current QSTDI approach demonstrates estimation of just one relevant fire hazard property, but may similarly be applied to develop estimation methods for others. The QSTDI approach facilitates fire hazard ranking and classification from molecular structure for compounds without available property data and, in particular, allows reliable identification of low- and high-fire-hazard chemicals.

6 CONCLUSIONS

Estimation of flash points from normal boiling points for Si-organic and Ge-organic compounds should rely on QPPRs especially developed for these compound classes. The QPPR for silanes of this work, the QPPR reported by Hsieh for silicones and the QPPR for germanes of this work allow reliable estimation of flash point. In case of missing boiling point data, QSTDs are proposed to estimate flash points of query compounds from flash points of structurally related compounds. An excellent correlation is found between flash points of germanes with flash points of silane analogues. Flash points of substituted silanes can be estimated by substituent exchange. Depending on the structure of the leaving and replacing substituent, a query flash point can either be estimated from a correlation equation or a boundary relation. The latter approach may not result in a highly accurate value, but typically gives a range within the value can be expected with high probability when a statistically significant number of compound pairs exhibiting the considered pattern of substituent exchange was employed. Thus, this approach suits flash point prediction very well, where, for example, labeling a query compound as "less flammable than a known low-flammability compound" or as "more flammable than a known high-flammability compound" is more beneficial to fire hazard assessment than deriving a highly accurate numerical flash point value.

REFERENCES

- Arkles, B. (Editor), *Silicon, Germanium, Tin and Lead Compounds, Metal Alkoxides, Diketonates and Carboxylates. A Survey of Properties and Chemistry*. Gelest, Inc., Tullytown, PA 19007, 2000.
- Catoire, L., and Naudet, V, A unique equation to estimate flash points of selected pure liquids. *J. Phys. Chem. Ref. Data*, 33:1083–1111, 2004.
- Drefahl, A., The list of flash point and boiling point data selected from Arkles [2000] for this study are available from the author upon request, 2006.
- Drefahl, A., and Reinhard, M, Similarity-based search and evaluation of environmentally relevant properties for organic compounds in combination with the group contribution approach. *J. Chem. Inf. Comput. Sci.*, 33:886–895, 1993.
- Hsieh, F.-Y., Note: Correlation of closed-cup flash points with normal boiling points for silicone and

general organic compounds. *Fire Mater.*, 21:277–282, 1997.

Katritzky, A. R., et al, QSPR analysis of flash points. *J. Chem. Inf. Comput. Sci.*, 41:1521–1530, 2001.

Muruga, R., et al, Predicting physical properties from molecular structure. *CHEMTECH*, pages 17–23, June 1994.

Paralikas, A. N., and Lygeros, A. I., A multi-criteria and fuzzy logic based methodology for the relative ranking of the fire hazard of chemical substances and installations. *Trans IChemE, Part B, Process Safety and Environmental Protection*, 83(B2):122–134, 2005.

Reinhard, M., and Drefahl, A., *Handbook for Estimating Physicochemical Properties of Organic Compounds*. John Wiley & Sons, Inc., New York, 1999.

Spencer, A. B., and Colonna, G. R. (Editors), *Fire Protection Guide To Hazardous Materials. 13TH Edition*. National Fire Protection Association, Quincy, MA 02269, 2002.

Tetteh, J., et al, Quantitative structure-property relationships for the estimation of boiling point and flash point using a radial basis function neural network. *J. Chem. Inf. Comput. Sci.*, 39:491–507, 1999.