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

Axel Drefahl

axel@owenstechnology.com

Overview

- Definitions and data sets

- Normal boiling point (T_{nb}) - flash point (T_f) correlations

- Molecular similarity/difference approach for flash point modeling

- Database integration of molecular difference methods

- Conclusions

Definition of flash point

The flash point of a substance is the minimum **temperature** at which sufficient vapor is emitted to form an ignitable mixture with air near the surface of the substance inside a testing apparatus:

- Open-cup flash points, $T_{\mathsf{f,oc}}$

- Closed-cup flash points,
  $T_{\mathsf{f,cc}} \approx T_{\mathsf{f,oc}} + 5°\mathsf{C}$

Typical measured values:

- Low value: $T_{\mathsf{f}} = -38\,°\mathsf{C}$ for acetaldehyde

- High value: $T_{\mathsf{f}} = 232\,°\mathsf{C}$ for diisooctyl phthalate

Fire hazard rating

4 - Gases and liquids with $T_f/°C < 22.8$ and $T_{nb}/°C < 37.8$

3 - Liquids with $T_f/°C < 22.8$ and $T_{nb}/°C > 37.8$ or $22.8 < T_f/°C < 37.8$

2 - Liquids with $37.8 < T_f/°C < 93.4$

1 - Liquids and solids with $T_f/°C > 93.4$

0 - Materials that will not burn (in air exposed to 815.5°C for 5 min.)

For further details see, for example:
`http://www.ehs.neu.edu/flammabi.htm`
or National Fire Protection Association (NFPA) references

Data set $(T_{f,cc}, T_{nb})$

- Silanes, $SiR_1R_2R_3R_4$ $(n = 123)$:
  $-27 \leq T_f/^\circ C \leq 175$;
  $36 \leq T_{nb}/^\circ C \leq 304$;
  $1 \leq N_C \leq 16$

- Germanes, $GeR_1R_2R_3R_4)$ $(n = 13)$:
  $-19 \leq T_f/^\circ C \leq 160$;
  $43 \leq T_{nb}/^\circ C \leq 274$;
  $1 \leq N_C \leq 16$

where R is H, Cl, Br, alkyl, alkenyl, phenyl, alkoxy, acetyl, or acetoxy and N_C = number of carbon atoms per molecule.

Data source: *Silicon, Germanium, Tin, and Lead Compounds, Metal Alkoxides, Diketonates and Carboxylates. A Survey of Properties and Chemistry.* Edited by B. Arkles. Gelest, Inc., Tullytown, PA 19007, 2000.

Quantitative T_{nb}/T_f Relationships

- Catoire and Naudet (T in K) with organic compounds:

$$T_f = 0.354 \cdot (T_{nb}^{1.14711}) \cdot (N_C^{-0.768})$$

- Hsieh (T in °C) with silicones:

$$T_f = -51.24 + 0.45 T_{nb} + 0.0005 T_{nb}^2$$

- This work with silanes and germanes:

$$T_f = a_0 + a_1 T_{nb} + a_2 N_C$$

QPPR Results

Quantitative T_{nb}/T_f Relationship for silanes:

$$T_f = -59.830 + 0.707 T_{nb} - 1.319 N_C$$

$$n = 86, r = 0.965, F = 556.6$$

Quantitative T_{nb}/T_f Relationship for germanes:

$$T_f = -42.918 + 0.571 T_{nb} - 1.378 N_C$$

$$n = 12, r = 0.953, F = 44.33$$

Molecular similarity/difference approach

Main features:

- Use of reference/database compounds with measured T_f

- Only structure-input in estimation request

- Quantitative property estimation

- Qualitative estimation: lower and upper property limit

Expected results of estimation request:

- Prioritized list of estimates

- Explained estimates (worked-example style)

Quantitative source/target difference (QSTD) are based on formal source/target transformation (STT): $S \rightarrow T$

STTs define binary ordered subsets of a set of molecular graphs.

Example 1: Formal replacement of Si by Ge

S=Trichlorosilane: $SiHCl_3$

T=Trichlorogermane: $GeHCl_3$

Example 2: Substitution of methyl by iso-propyl group

S=Dichloromethylsilane: $SiCl_3CH_3$

T=Dichloroisopropylsilane: $SiCl_3CH(CH_3)_2$

Working hypothesis: A particular SST is associated with a functional expression of the change in property value.

Quantitative silane/germane relationship

Evaluation data:

STT: $S_i \rightarrow T_i$	$T_f\{S_i\}$	$T_f\{T_i\}$	D_i
SiCl$_2$Ph$_2$ $\rightarrow$ **Ge**Cl$_2$Ph$_2$	157	160	3
$\cdots$	$\cdots$	$\cdots$	$\cdots$
SiMe$_3$Ph $\rightarrow$ **Ge**Me$_3$Ph	40	53	13
$\cdots$	$\cdots$	$\cdots$	$\cdots$
SiCl$_2$Me$_2$ $\rightarrow$ **Ge**Cl$_2$Me$_2$	-10	21	31

$T_f\{S_i\}$ and $T_f\{T_i\}$ in °C;
$D_i = T_f\{T_i\} - T_f\{S_i\}$;
$i = 1, \ldots, m$ with $m =$ number of source/target pairs.

QSTD:
$T_f\{T_i\} = 19.047 + 0.915\,T_f\{S_i\}$
$m = 13, \ \ r = 0.983$

QSTD Interval (QSTDI):
$B_l^{\triangle} = 3$ to $B_u^{\triangle} = 31$

Definition of QSTD Interval (QSTDI)

Empirical lower and upper STT boundary:

$B_l^{\triangle} = \min\{D_i \mid 1 \le i \le m\}$
$B_u^{\triangle} = \max\{D_i \mid 1 \le i \le m\}$
where $D_i = T_{\mathsf{f}}\{\mathsf{T}_i\} - T_{\mathsf{f}}\{\mathsf{S}_i\}$
and $m =$ number of source/target pairs.

Interval estimate for query compound, $\mathbf{Q} = \mathsf{T}_k$, from source S_k ($k > m$) when STT is given by $\mathsf{S}_k \to \mathsf{Q}$:

$$T_{\mathsf{f}}\{\mathsf{S}_k\} + B_l^{\triangle} \le T_{\mathsf{f}}\{\mathbf{Q}\} \le T_{\mathsf{f}}\{\mathsf{S}_k\} + B_u^{\triangle}$$

Quantitative methylsilane/benzylsilane relationship

Evaluation data:

SST: $S_i \rightarrow T_i$	$T_{a,i}$	$T_{b,i}$	D_i
SiMe$_4$ $\rightarrow$ SiMe$_3$Bz	-27	62	89
SiMeEx$_3$ $\rightarrow$ SiBzEx$_3$	30	127	97
SiClMe$_3$ $\rightarrow$ SiClMe$_2$Bz	-27	73	100
SiCl$_3$Me $\rightarrow$ SiCl$_3$Bz	-15	87	102
SiHMe$_3$ $\rightarrow$ SiHMe$_2$Bz	< 20	65	> 85

Me=methyl, Ex=ethoxy, Bz=benzyl;
$T_{a,i} = T_f\{S_i\}$; $T_{b,i} = T_f\{T_i\}$;
$D_i = T_f\{T_i\} - T_f\{S_i\}$.

QSTD:
$T_f\{T_i\} = 97.27 + 1.027\,T_f\{S_i\}$
$m = 4, \ r = 0.980$

QSTDI:
$B_l^{\triangle} = 89$ to $B_u^{\triangle} = 102$

Change of fire hazard rating for methyl-to-benzyl transformation in silanes

Fire hazard ratings:

SST: $S_i \rightarrow T_i$	R$\{S_i\}$	R$\{T_i\}$	D_i
$SiMe_4 \rightarrow SiMe_3Bz$	4	2	-2
$SiMeEx_3 \rightarrow SiBzEx_3$	3	1	-2
$SiClMe_3 \rightarrow SiClMe_2Bz$	4	2	-2
$SiCl_3Me \rightarrow SiCl_3Bz$	4	2	-2
$SiHMe_3 \rightarrow SiHMe_2Bz$	4	2	-2

Me=methyl, Ex=ethoxy, Bz=benzyl;

R = fire hazard rating;

D_i = R$\{T_i\}$ − R$\{S_i\}$.

Comparison with group contribution models (GCMs)

Source/ target transformations (STTs) are group-group exchanges, hence STTs ($\rightarrow$ QSTDs) provide excellent methods to test validity and accuracy of group contribution models (GCMs).

Since GCMs assume group additivity for a property P, the following should be true for any given source/target set:

$$D_i = P_i\{\mathsf{T}_i\} - P_i\{\mathsf{S}_i\} = \text{constant}$$

and

$$B_u^\triangle - B_l^\triangle = 0$$

Current software-approach in QSTD analysis

- Use of in-house databases

- Structure encoding with Simplified Molecular Input Line Entry System (SMILES) notation;
  for example, dimethoxyphenylgermane:

  ```
  c1ccccc1[GeH](OC)OC
  ```

- Programming in Python

- Selecting source/target pairs by automatic recognition of molecular formula differences + human interaction

Graph isomorphism problem:

How to prove that a pair of structures comply with a given SST?

Future software for QSTD analysis and applications?

- Convert source SMILES to target SMILES (or vice versa)

  1. Locate exchanging group in source SMILES

  2. Replace in compliance with given STT to derive target SMILES

  3. Get unique target SMILES by applying CANGEN-algorithm (CANonicalization to GENerate unique notation)

  4. Apply unique target SMILES as search key

- Is chemical mark-up language (CML) applicable to Internet source/target querying?

- Any comments and suggestions are welcome!

Conclusions

- As for organic compounds, the boiling point is the key parameter to estimate flash points of silanes and germanes. (Note: $QT_{nb}T_f$Rs limited since boiling point often at reduced temperature.)

- Source/target transformations (STTs) describe molecular (dis)similarity and provide a general tool to relate new structures (queries) to existing ones in databases.

- Quantitative source/target difference (QSTD) relations allow molecular-structure-based flash point estimation(s) for a query from known flash point data of database compounds.

- Integration of QSTDs with databases and on-line services needs future evaluation and exploration!