AIChE

Property Prediction and Consistency Analysis by a Reference Series Method

Mordechai Shacham and Inga Paster
Dept. of Chemical Engineering, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel

Neima Brauner
School of Engineering, Tel-Aviv University, Tel-Aviv 69978, Israel

DOI 10.1002/aic.13813

Data analysis and prediction of pure component properties of long-chain substances is considered. The emphasis is on homologous series and properties for which insufficient data are available. A two-stage procedure is recommended, whereby a linear (or nonlinear) quantitative structure–property relationship (QSPR) is fitted to a "reference" series, for which an adequate amount of precise data is available. This QSPR should represent correctly both the available data and the asymptotic behavior of the property. In the second stage a quantitative property-property relationship (QPPR) is derived to represent the predicted property values of a "target" series in terms of the property values of the reference series. The procedure is applied for properties which are highly correlated with the number methylene groups in homologous series: ΔH_f^0 *and* ΔS_f^0*. It is shown that the method is very useful for consistency analysis of property data and enables a reliable prediction of* ΔH_f^0 *and* ΔS_f^0*, and, thus, also of* ΔG_f^0 *for long-chain substances.* © 2012 American Institute of Chemical Engineers *AIChE J,* 59: 420–428, 2013

Keywords: property prediction, homologous series, QSPR, reference series, QPPR

Introduction

Long-chain substances pose special challenges for property prediction, as many of their properties cannot be measured because of thermal instability. Asymptotic behavior correlations (ABCs, e.g., Marano and Holder,[1] Nikitin et al.[2]) are currently widely used for property prediction of long-chain substances. The ABCs are nonlinear correlations in terms of the number of carbon atoms (n_C), which are derived using available experimental data for low n_C compounds and an estimation of the property value at the limit $n_C \rightarrow \infty$, y^{∞}. The use of nonlinear correlations with $\sim$ 6 adjustable parameters for extrapolation (as done with ABCs) can be very risky, in particular for homologs and properties for which a small amount of and/or inaccurate data are available.

Recently Paster et al.[3] presented a technique for applying linear quantitative structure property relationships (QSPRs) for the prediction of properties of long-chain substances in homologous series. Using this method, molecular descriptors collinear with a particular property are identified based on available experimental data. From among these, the descriptors whose asymptotic behavior is similar to the property behavior are eventually used for prediction. If no descriptor that adequately satisfies both conditions can be found, a nonlinear correction term is added to the QSPR.

The use of the method of Paster et al.[3] can be challenging for series and properties for which an insufficient amount of and/or low-precision data are available. To obtain reliable predictions in such cases, the "Reference Series" method has been developed (Paster et al.[4] Paster et al.[5]), where the QSPR is derived only for one "reference" homologous series for which the largest amount of and highest precision experimental data are available. Other homologous series (target series) are related to the reference series by means of a quantitative property-property relationship (QPPR). This method enables assessing the consistency of the experimental data available for the target series and increases the prediction reliability when long-range extrapolation is involved.

While testing the reference series method we have found that for some properties, like enthalpy of formation of ideal gas at 298.15 K: ΔH_f^0, entropy of formation of ideal gas at 298.15 K: ΔS_f^0 and Gibbs energy of formation of ideal gas at 298.15 K: ΔG_f^0, linear QSPRs and QPPRs can be fitted as function of the number of carbon atoms (n_C). For other properties, often different descriptors (instead of n_C) should be used in the QSPR and/or nonlinear functions need to be used in both the QSPR and QPPR to adequately represent the property.

In this article, the reasons for the different behavior of the properties in homologous series are investigated, and the benefits of the use of linear QSPRs and QPPRs for analyzing data consistency and making predictions involving short- and long-range extrapolation are demonstrated.

Basic Principles and an Introductory Example

Homologous series have the general formula $H(CH_2)_nR$, where $-CH_2-$ is the methylene group, n is the number of methylene groups, and R is an end (or functional) group,

---

Correspondence concerning this article should be addressed to M. Shacham at shacham@bgu.ac.il.

Table 1. Enthalpy of Formation for Ideal Gas Data for the *n*-alkane and 1-alkene Homologous Series

		ΔH_f^{0}[1]			ΔH_f^{0}[1]	
Component Name	No. of C-atoms	Value[2] (kJ/mol)	Uncertainty	Component Name	Value[2] (kJ/mol)	Uncertainty
methane	1	**-74.52**	<1%			
ethane	2	**-83.82**	<1%	ethylene	**52.51**	< 3%
propane	3	**-104.68**	< 1%	propylene	***20.23***	< 1%
n-butane	4	**-125.79**	< 1%	1-butene	***-0.50***	< 3%
n-pentane	5	**-146.76**	< 1%	1-pentene	***-21.62***	< 3%
n-hexane	6	**-166.94**	< 1%	1-hexene	***-41.67***	< 3%
n-heptane	7	**-187.65**	< 1%	1-heptene	**-62.89**	< 3%
n-octane	8	**-208.75**	< 1%	1-octene	***-81.94***	< 3%
n-nonane	9	**-228.74**	< 1%	1-nonene	***-103.50***	< 3%
n-decane	10	**-249.46**	< 1%	1-decene	**-124.70**	< 3%
n-undecane	11	**-270.43**	< 1%	1-undecene	***-144.90***	< 3%
n-dodecane	12	**-290.72**	< 1%	1-dodecene	***-165.40***	< 1%
n-tridecane	13	**-311.77**	< 1%	1-tridecene	-186.20	< 3%
n-tetradecane	14	**-332.44**	< 1%	1-tetradecene	-206.80	< 3%
n-pentadecane	15	**-353.11**	< 1%	1-pentadecene	-228.00	< 3%
n-hexadecane	16	**-374.17**	< 1%	1-hexadecene	**-249.59**	< 1%
n-heptadecane	17	**-394.45**	< 1%	1-heptadecene	-269.30	< 3%
n-octadecane	18	**-415.12**	< 1%	1-octadecene	-290.00	< 3%
n-nonadecane	19	**-435.79**	< 1%	1-nonadecene	-310.70	< 3%
n-eicosane	20	**-456.46**	< 1%	1-eicosene	-331.30	< 3%

[1]Source: DIPPR database[9]
[2]Experimental data are shown in bold, Smoothed data are shown in bold-italic

which defines the homologous series. We consider the change of properties as function of *n*.

It is generally accepted (see for example Marano and Holder[6]) that some properties (e.g., liquid molar volume, ρ_L; enthalpy of formation of ideal gas, and critical volume V_C) are additive in nature with a monotonic incremental change with increasing *n* (a type II property). Other properties, such as the normal melting (T_m) and boiling (T_b) points, and the critical temperature (T_C) approach a finite value (y^{∞}) for large carbon numbers, $n \to \infty$ (a type I property). Upon approaching the limit, the difference between the property values of the various series monotonically decreases, as the effect of specific functional groups diminishes. Thus, a particular property should approach the same limiting value for different series.

Based on the properties of the homologous series, Paster et al.[3,4] suggested the use of a linear QSPR for predicting properties of long-chain substances: using the available experimental data, molecular descriptors which are collinear with the property are identified, whereby

$$y = \beta_0 + \beta_1 \zeta^* \tag{1}$$

where ζ^* is a molecular descriptor which is collinear with the property in the range where experimental data are available, y is the property and β_0 and β_1 are parameters obtained by regression of the experimental data. In case of a property of type I, an additional requirement regarding the selected descriptor ζ^* is that the resultant linear QSPR converges to a generally accepted y^{∞} value.

Peterson[7,8] has recently shown that properties of compounds in two families of compounds of the same kind (e.g., homologous series) can be represented (at least locally) by a linear QPPR

$$y_t = \mathrm{B}_0 + \mathrm{B}_1 y_r \tag{2}$$

where y_r is the property value of a compound in the reference series, y_t is the property of a compound (with the same *n* or n_C) in the target series and B_0 and B_1 are parameters obtained by regression of the experimental data.

Let us consider the change of the property $y = \Delta H_f^0$ for the *n*-alkane and 1-alkene series. The recommended ΔH_f^0 values for the first 20 members of these series from the DIPPR (Rowley et al.[9]) database are shown in Table 1. Observe that there are three types of recommended values: experimental—that usually means that there are several experimental data points available in the literature and one of them was selected as the "recommended" value based on the analysis of the DIPPR staff; smoothed—the same as experimental, except that the recommended value does not exactly match any of the experimental ΔH_f^0 values (as in the case of an average value); and predicted—when no experimental data are available, a predicted value is recommended. In the case of the *n*-alkane series, experimental ΔH_f^0 values are available for all the first 20 members shown in Table 1. For the 1-alkene series, experimental or smoothed values are available for 12 members (mostly for lower carbon numbers) and predicted values for 7 compounds of higher carbon numbers.

"Uncertainties" (upper limits on experimental or prediction errors) are also reported in the DIPPR database. The uncertainties of all the ΔH_f^0 values of the *n*-alkane series are <1% (Table 1). For the 1-alkene series the uncertainties are higher: they are <1% only for three compounds and <3% for the rest.

Let us denote the difference between the ΔH_f^0 values of the members of the series with $n_C = i + 1$ and $n_C = i$ as $\delta y_{(i+1,i)} = \Delta H^0_{f(i+1)} - \Delta H^0_{f(i)}$. Relying on several earlier publications, Sellers et al.[10] suggested that starting at certain $n_C \geq n_{C,\min}$ the value of $\delta y_{(i+1,i)}$ is constant, independent of the value of n_C. Using the *n*-alkane data of Table 1, $\delta y_{(2,1)} = -9.3$ kJ/mol; and $\delta y_{(3,2)} = -20.86$ kJ/mol. The average of the $\delta y_{(i+1,i)}$ for the rest of the compounds is $\delta \bar{y}_{(i+1,i)} = -20.7$ kJ/mol with standard deviation of $\sigma = 0.33$ kJ/mol. Using these results Eq. 1 can be rewritten for ΔH_f^0 of the *n*-alkane series as

$$y = \beta_0 + n_C \delta \bar{y}_{(i+1,i)} \quad n_C \geq n_{C,\min} \tag{3}$$

where $\delta \bar{y}_{(i+1,i)} = -20.72$ kJ/mol, and $\beta_0 = -104.68 - 3(-20.7) = -42.58$ kJ/mol, and $n_{C,\min} = 3$. These results are consistent with the results of Sellers et al.[10] who found the values $\delta\bar{y}_{(i+1,i)} = -20.72$ kJ/mol, and $\beta_0 = -42.65$ kJ/mol.

The first compound belonging to the 1-alkene series is ethylene, with $n_C = 2$. Using the data of Table 1, $\delta y_{(3,2)} = -32.28$ kJ/mol; $\delta y_{(4,3)} = -20.73$ kJ/mol, and the average for the rest of the compounds is $\delta\bar{y}_{(i+1,i)} = -20.68$ kJ/mol with standard deviation of $\sigma = 0.66$ kJ/mol. This is essentially the same value as for the *n*-alkane series. Thus, the incremental change of ΔH_f^0 with n_C is constant, and it is of the same value for various series provided that n_C is greater than some $n_{C,\min}$. Hence, in the QPPR model (Eq. 2), $B_1 = 1$. Combining Eqs. 3 and 2 yields

$$y_t = B_0 + y_r = B_0 + \beta_0 + n_C\delta\bar{y}_{(i+1,i)} \quad n_C \geq n_{C,\min} \tag{4}$$

For the 1-alkene series $n_{C,\min} = 3$. The value of B_0 is obtained from Eq. 5

$$B_0 = \delta y_{tr} = y_t - y_r = \Delta H_{ft}^0 - \Delta H_{fr}^0 \quad n_C \geq n_{C,\min} \tag{5}$$

Preferably an average value $\delta\bar{y}_{tr}$ should be used, which is obtained by considering all the available experimental data for the two series with corresponding $n_C > n_{C,\min}$.

The use of the reference series technique is especially important for series where the amount and/or precision of the experimental data are insufficient to determine the slope of the QSPR and of the QPPR. In such cases, predictions which involve extrapolation to high n_C values may be highly inaccurate. However, ΔH_f^0, ΔS_f^0 and ΔG_f^0 are type II properties of the gaseous phase which are defined at a constant temperature and pressure. In this case, the incremental change of the property value with n_C ($> n_{C,\min}$) results from the addition of $-CH_2-$ group both in the reference series and in the target series. Therefore, the slope of the QPPR model (B_1 in Eq. 2) can be set to 1 based on physical considerations, whereas the data is used only for deducing the appropriate value of the parameter B_0. Obviously, for type II properties, the contribution of B_0 to the predicted value diminishes when extrapolating to high n_C. Hence, high-precision QPPR which allow a long-range interpolation can be derived based on a limited data of the property value of the target series.

In the remainder of the article practical approaches for using Eqs. 3, 4 and 5 for modeling ΔH_f^0, ΔS_f^0 and $\Delta G_f^0 = \Delta H_f^0 + T\Delta S_f^0$ at $T = 298.15$ K, and practical use of these equations for analysis of the consistency of the data will be demonstrated.

Methodology

To carry out the studies reported here our physical property and molecular descriptor database was used. This database contains physical property data for 1,798 compounds. For each compound there are numerical values and data uncertainty for 34 constant properties (critical properties, normal melting and boiling temperatures, heat of formation, flammability limits etc.). All the property data are from the DIPPR database (Rowley et al.[9]). The database also contains 3,224 molecular descriptors for the 1,798 compounds, which were generated by the Dragon, version 5.5. Software (Todeschini and Consonni[11]) from minimum energy 3-D molecular models. The molecular structure (mol) files were provided by Rowley.[12]

The *n*-alkane series, for which the largest amount of experimental data is available, can serve as a reference series (see for example, Greenshields and Rossini[13]). To derive a QSPR for the reference series, the available property (experimental) data was divided into a training set (typically 10 compounds, unless a smaller number of data points were available) and a validation set. In cases where less than 10 experimental data points were available, predicted values reported by DIPPR were used in the validation set. The property values of the validation set compounds were used to verify the correctness of the reference series QSPRs. All the available experimental data (excluding outliers) were used for deriving QPPRs.

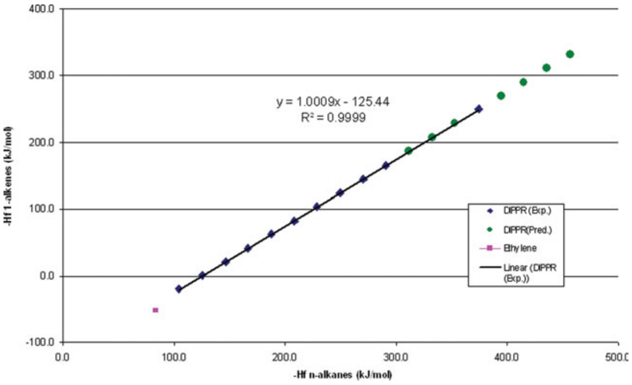

Figure 1. Plot of ($-\Delta H_{ft}^0$) vs. ($-\Delta H_{fr}^0$) for the 1-alkene and *n*-alkane series.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Calculations of the regression model parameters were carried out using the linear and nonlinear regression programs of the Polymath package (Polymath is a product of Polymath Software, http://www.polymath-software.com).

A QPPR for the Heat of Formation of the *n*-alkane (Reference) and 1-alkene (Target) Series

Let us consider the relationship between the enthalpy of formation of ideal gas at 298.15 K for the *n*-alkane and the 1-alkene series. The heat of formation of the 1-alkene (target) series is denoted as ΔH_{ft}^0 and that of the *n*-alkane (reference) series by ΔH_{fr}^0.

Plotting ($-\Delta H_{ft}^0$) vs. ($-\Delta H_{fr}^0$), using the data from Table 1, for compounds with the same n_C (Fig. 1) yields a straight line with a slope of $B_1 = 1.0009 \pm 0.003$, a free parameter $B_0 = -125.44 \pm 0.7$ and a correlation coefficient $R^2 = 0.9999$ (only experimental and smoothed data points were used for regression). These results confirm the expectation that the heat of formation of the 1-alkene series can be represented by a linear QPPR (Eq 3) with $y_t = \Delta H_{ft}^0$, $y_r = \Delta H_{fr}^0$ and $B_1 = 1$ (as argued in the ''Basic Principles'' section).

Table 2 shows results of data analysis when values of ΔH_f^0 for 18 homologs of matching n_C from the two series are examined (compounds with $n_C < n_{C\min} = 3$ are not included in the comparison). The difference $\delta y_{tr} = y_t - y_r = \Delta H_{ft}^0 - \Delta H_{fr}^0$ is shown in the 2nd column of the table. Observe that all the differences are in the 124.5–125.7 (kJ/mol) range, except for $n_C = 8$ (1-octene). For 1-octene, the DIPPR recommended value is $\Delta H_f^0 = -81.94$ kJ/mol, but several other experimental values are available, as well. The experimental value reported by Sakano et al.[14] is $\Delta H_f^0 = -83.6$ kJ/mol, whereby $\delta y_{tr} = -125.15$ kJ/mol is within the range set by the rest of the compounds.

The δy_{tr} values are plotted in Figure 2 vs. n_C. The values are randomly distributed around $\delta y_{tr} \approx -125.15$ kJ/mol and 1-octene ($n_C = 8$, using the DIPPR recommended value for

Table 2. The Relationship Between ΔH_f^0 of the *n*-alkane (Reference) and 1-alkene (Target) Series

No. of C-atoms	$\delta y_{tr} = \Delta H_{ft}^0 - \Delta H_{fr}^0$	ΔH_f^0 (predicted) Value (J/kmol)	Residual	Prediction Error (%)[1]
3	124.91	20.56	-0.33	**-1.62**
4	125.29	-0.55	0.05	**-10.55**
5	125.14	-21.52	-0.10	0.45
6	125.27	-41.70	0.03	-0.08
7	124.76	-62.41	-0.48	0.76
8	126.81	-83.51	1.57	**-1.92**
9	125.24	-103.50	0.00	0.00
10	124.76	-124.22	-0.48	0.38
11	125.53	-145.19	0.29	-0.20
12	125.32	-165.48	0.08	-0.05
13	125.57	-186.53	0.33	-0.18
14	125.64	-207.20	0.40	-0.19
15	125.11	-227.87	-0.13	0.06
16	124.58	-248.93	-0.66	0.26
17	125.15	-269.21	-0.09	0.03
18	125.12	-289.88	-0.12	0.04
19	125.09	-310.55	-0.15	0.05
20	125.16	-331.22	-0.08	0.02
Average	125.24			
St. Dev.	0.60			

[1] Prediction error higher than data uncertainty are shown in bold.

ΔH_f^0) can be clearly identified as an outlier. In fact, a graphical representation of δy_{tr} can be used for consistency analysis of the available heat of formation data. In this case, the distribution of the DIPPR predicted values is similar to the distribution of the experimental values, indicating that the predicted values are consistent with the available experimental data for these two series.

The average δy_{tr} value (excluding 1-octene), as calculated from the experimental data (Table 2), is 125.24 kJ/mol with a standard deviation of 0.6 kJ/mol. This average can be used in the QPPR for as an estimate for B_0, (i.e., $\Delta H_{fr}^0 = \Delta H_{ft}^0 + 125.24$). The ΔH_{ft}^0 values, thus, predicted, the residuals (deviations from the DIPPR reported values, ε) and the prediction error (%) are also shown in Table 2. Excluding 1-octene from consideration, the residuals are in a narrow range of $0.01 \leq \text{abs}(\varepsilon) \leq 0.49$ kJ/mol. The variability in the prediction error (%) is much higher than that because of the almost four-orders of magnitude increase in the absolute values between low and high carbon number compounds. The prediction error is higher than the data uncertainty for propylene (uncertainty <1%, prediction error 1.672%) and 1-butene (uncertainty <3%, prediction error 10.55%). For propylene, using the value recommended by Luria and Benson[15] of 20.5 kJ/mol, rather than the DIPPR recommended value, reduces the prediction error to 0.37%. For 1-butene, four experimental data points, in the range of $-0.54 \leq \Delta H_f^0 \leq -0.1$ ranges, with uncertainties of up to 100%, are available in the literature. These imply a higher uncertainty than that assigned to the DIPPR recommended value for 1-butene (< 3%). It is apparent that for such a small absolute values, the data uncertainty is rather high.

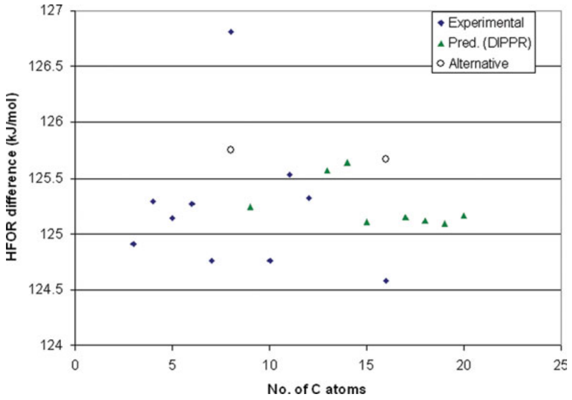

Figure 2. Plot of $\delta y_{tr} = \Delta H_{ft}^0 - \Delta H_{fr}^0$ for the *n*-alkane (reference) and 1-alkene (target) series vs. n_C.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

A QPPR for ΔH_f^0 of the *n*-alkane (Reference) and 1-alcohol (Target) Series—Effects of Data Smoothing

Data related to ΔH_f^0 of the members of the 1-alcohol series in the range of $1 \leq n_C \leq 20$ are shown in Table 3. The data include the DIPPR recommended ΔH_f^0 values, as well as the minimal and maximal reported experimental values. For many of the compounds the difference between the maximal and minimal values is smaller than the uncertainty of the recommended value of 1%. However, there are also cases where the differences are considerably higher (7.3% for 1-nonanol, 2.1% for 1-hexadecanol, etc). Most of the DIPPR recommended data is categorized as smoothed. The values of δy_{tr} for the 1-alcohol series shown in Table 4 for $n_C \geq n_{C\min} = 2$ are in the range of: $-151.2 < \delta y_{tr} < -147.7$ kJ/mol.

Figure 3 shows the δy_{tr} values when plotted vs. n_C. In this case there is a clear trend of increasing δy_{tr} values with increasing n_C. To explain this trend, the data is used to fit a linear QPPR ($\Delta H_{ft}^0 = B_0 + B_1 H_{fr}^0$) to the 15 compounds in the range of $2 \leq n_C \leq 16$. The parameter values obtained are $B_0 = (-150.7286 \pm 0.66)$ and $B_1 = (0.9922251 \pm 0.0027)$ with a correlation coefficient $R^2 = 0.99998$. Thus, there is a small yet significant deviation of B_1 from the expected value of one. The irregular behavior of the 1-alcohol series data was

Table 3. Enthalpy of Formation for Ideal Gas Data for the 1-alcohol Homologous Series

Component Name	No. of C-atoms	ΔH_f^0[1] Value[2] (kJ/mol)	Uncertainty	min. value (kJ/mol)	max. value (kJ/mol)
methanol	1	**-200.94**	<1%	**-201.20**	**-200.66**
ethanol	2	**-234.95**	<1%	**-235.10**	**-234.40**
1-propanol	3	***-254.60***	< 1%	**-257.50**	**-255.00**
1-butanol	4	***-275.10***	< 1%	**-275.30**	**-274.40**
1-pentanol	5	***-295.70***	< 1%	**-302.40**	**-295.60**
1-hexanol	6	***-316.20***	< 1%	**-318.00**	**-315.10**
1-heptanol	7	***-336.80***	< 1%	**-341.60**	**-330.90**
1-octanol	8	***-357.30***	< 1%	**-381.20**	**-353.00**
1-nonanol	9	***-377.90***	< 1%	**-382.00**	**-375.60**
1-decanol	10	***-398.50***	< 1%	**-403.30**	**-394.00**
1-undecanol	11	***-419.00***	< 1%	-	-
1-dodecanol	12	***-439.60***	< 1%	**-439.70**	**-436.50**
1-tridecanol	13	***-460.10***	< 1%	-	-
1-tetradecanol	14	***-480.70***	< 1%	**-484.00**	**-476.00**
1-pentadecanol	15	***-501.20***	< 1%	**-504.70**	**-501.80**
1-hexadecanol	16	***-521.80***	< 1%	**-525.30**	**-514.20**
1-heptadecanol	17	-542.30	< 1%	-	-
1-octadecanol	18	-562.90	< 1%	-	-
1-nonadecanol	19	-583.50	< 1%	-	-
1-eicosanol	20	-605.20	< 3%	-	-

[1] Source: DIPPR database[9]

[2] Experimental data are shown in bold, Smoothed data are shown in bold-italic

Table 4. The Relationship Between ΔH_f^0 of the *n*-alkane and 1-alcohol (Target) Series

No. of C-atoms	$\delta y_{tr} = \Delta H_{ft}^0 - \Delta H_{fr}^0$	ΔH_f^0 (predicted) Value (J/kmol)	Residual	Prediction Error (%)
2	-151.13	-232.61	-2.34	1.00
3	-149.92	-253.47	-1.13	0.44
4	-149.31	-274.58	-0.52	0.19
5	-148.94	-295.55	-0.15	0.05
6	-149.26	-315.73	-0.47	0.15
7	-149.15	-336.44	-0.36	0.11
8	-148.55	-357.54	0.24	-0.07
9	-149.16	-377.53	-0.37	0.10
10	-149.04	-398.25	-0.25	0.06
11	-148.57	-419.22	0.22	-0.05
12	-148.88	-439.51	-0.09	0.02
13	-148.33	-460.56	0.46	-0.10
14	-148.26	-481.23	0.53	-0.11
15	-148.09	-501.90	0.70	-0.14
16	-147.63	-522.96	1.16	-0.22
17	-147.85	-543.24	0.94	-0.17
18	-147.78	-563.91	1.01	-0.18
19	-147.71	-584.58	1.08	-0.19
20	-148.74	-605.25	0.05	-0.01
Average	-148.79			
St. Dev.	0.59			

also noticed by Sellers et al.[10] but they did not provide satisfactory explanation to the discrepancy.

It can be argued however, that in light of the wide spread of the available experimental data (see Figure 4) some deviation of the smoothed data from the theoretical trend is possible. The proposed method enables detection of such deviations. Imposing consistency with the reference series (*n*-alkane) data, namely $B_1 = 1$, with $B_0 = -148.79$ (the average δy_{tr} value) yields, for the QPPR, $\Delta H_{ft}^0 = \Delta H_{fr}^0 - 148.79$. The median prediction error (based on absolute deviation from the DIPPR recommended values) is 0.11%, and the maximal error is 1%, thus, all predicted values are within the data uncertainty range.

Summary of ΔH_f^0-QPPRs for 12 homologous series

Table 5 presents a summary of the derivation of ΔH_f^0-QPPR for 12 homologous series: 1-alkene, *n*-alcohol, *n*-mercaptan, *n*-alkanoic acid, *n*-acetate, 1-alkyne, aldehyde, *n*-formate, *n*-amine, cis-2-alkene, trans-2-alkene and alkyl-benzene. In all cases the *n*-alkane series is used as reference series. The differences in the number of carbon atoms in the functional group of the various series R, results in different $n_{C\min}$ values ($= 1 +$ the number of carbon atoms in R), which are shown in the third column of Table 5. In two cases (aldehyde and 1-alkyne) the data of the compounds with $n_C = n_{C\min}$ still show significant influence of the functional end group, therefore, a higher n_C (shown within brackets) is used as lower limit for deriving the QPPR. The range of the heat of formation of the compounds from the various series at $n_C = n_{C\min}$ is $-44.5 \leq \Delta H_f^0 \leq 184.9$ kJ/mol. The range of the heat of formation for the highest n_C compounds for which data are available is $-806 \leq \Delta H_f^0 \leq 41$.

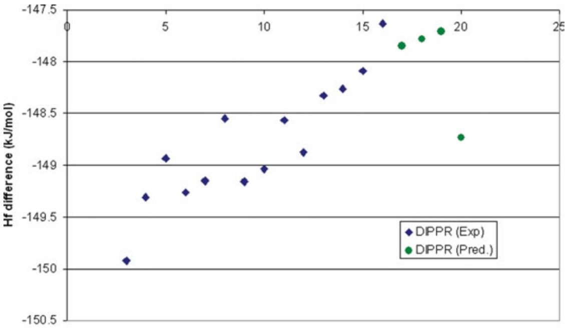

Figure 3. Plot of δy_{tr} for the *n*-alkane (reference) and 1-alcohol (target) series vs. n_C.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

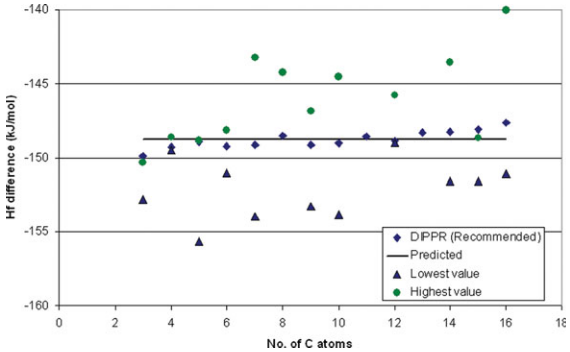

Figure 4. Plot of δy_{tr} for the *n*-alkane (reference) and 1-alcohol (target) series vs. n_C, including maximal, minimal and predicted values for the target series.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

The QPPR's were derived using experimental/smoothed data and were validated using both experimental and predicted data recommended by DIPPR. The number of experimental/smoothed data points available varied between 17 (for *n*-alkanoic acids) to 3 (*n*-formates and *n*-amines). In some cases the DIPPR recommended value for a compound was inconsistent with the rest of the values for the series and it had to be replaced by a different available experimental value. The list of these compounds is shown in Table 6. The list includes 7 compounds from the 1-alkene, *n*-alkanoic acid, cis-2-alkene and trans-2-alkene series. The sources of the alternative values are also shown in this Table 6.

The range of the QPPR B_0 values is -348 < B_0 < 238 kJ/mol, with standard deviations (σ) in the range of $0.25 \leq \sigma < 2.1$ kJ/mol (Table 5). Figure 5 shows the B_0 values of the various series when plotted vs. the ΔH_f^0 values of the compound with $n_C = 4$ (provided $n_{C,\min} \geq 4$). The B_0 values align along a straight line with a free parameter $a_0 = 125.51 \pm .55$, slope of $a_1 = 0.997 \pm .0.0023$ and correlation coefficient of $R^2 = 1$. Noting that for one value of n_C the value y_r (in Eq. 3) is constant, regardless of the target series and equal to ΔH_f^0 of *n*-butane ($-125.79 \approx$ a0), this plot suggests that the ΔH_f^0 values for the compounds involved are accurate enough to enable determination of B_0 based on a single data point, instead of averaging several ($\Delta H_{ft}^0 - \Delta H_{fr}^0$) values corresponding to $n_C \geq n_{C,\min}$.

The deviation (in %) between the QPPR predictions and the experimental/smoothed data are used to assess the accuracy of the representation of ΔH_f^0 by the QPPR. The median and the maximal values of these differences are reported in Table 5. The median in all the cases is < 1%, while the maximal deviation is larger than 1% in four series: 1-alkene, *n*-amine, trans-2-alkene and alkyl-benzene. In all cases the maximal deviation is associated with the smallest absolute ΔH_f^0 value.

Table 5. Summary of QPPR Derivation for ΔH_f^0 of 12 Series

			Range of ΔH_f^0 (used) kJ/mol		Data points[1]		QPPR		% Difference[2]	
No.	Series	n_{Cmin}[3]	at (lowest n_C)	at (highest n_C)	Exp./Smoothed	predicted	B_0	σ	Median	Maximal
1	1-alkene	3	20.23	−331.3(20)	11	7	125.25	0.48	0.18	8.56(-0.5)
2	*n*-alcohol	2	−234.96	−605.2(20)	14	4	−148.79	0.59	0.11	0.40
3	*n*-mercaptan	2	−46.3	−253.2(12)	11	-	37.76	0.52	0.26	0.86
4	*n*-alkanoic acid	2	−432	−806(20)	17	1	−347.79	2.08	0.23	0.67
5	*n*-acetate	4	−444.5	−609(12)	9	-	−318.26	0.37	0.07	0.1
6	1-alkyne	3(4)	184.9	41(10)	4	4	290.59	0.71	0.29	0.56
7	aldehyde	2(4)	−166.4	−394.1(13)	8	4	−81.80	0.48	0.13	0.47
8	*n*-formates	3	−388.3	−552(11)	3	6	−281.63	0.52	0.06	0.51
9	*n*-amines	2	−47.15	−298.2(14)	3	9	33.85	0.25	0.13	5.98(-47.15)
10	cis-2-alkenes	4	−7.4	−172.3(12)	5	3	118.44	0.44	0.36	0.9
11	trans-2-alkenes	4	−11	−342.2(20)	4	6	113.96	1.00	0.15	7.57(-11)
12	alkyl-benzenes	7	50.17	−302.5(24)	4	14	237.71	0.60	0.64	13.54(8.97)

[1] No. of compounds for which DIPPR recommended property values are available

[2] Differences between values predicted by the QPPR and the DIPPR recommended values. Respective heat of formation values are shown within brackets.

[3] The number in the brackets indicates the lowest n_C used for QPPR derivation when it differs from n_{Cmin}.

Prediction of ΔH_f^0 for the n-alkane series and the 12 target series

The ΔH_f^0 data for the *n*-alkane series (Table 1) can be used for derivation of a QSPR using the principles outlined in the "Basic Concepts" and "Methodology" sections. Ten compounds in the range of $10 \leq n_C \leq 19$ are used as the training set. As expected, one of the descriptors which have been identified as highly correlated with the ΔH_f^0 data is n_C. Fitting Eq. 1 (with $\zeta^* = n_C$ and $y = \Delta H_f{}^0$ kJ/mol) yielded the parameter values: $\beta_0 = (-42.52 \pm 0.7)$ and $\beta_1 = (-20.7 \pm 0.05)$ with a correlation coefficient $R^2 = 0.999991$. All the statistical indicators (confidence interval on the parameters, R^2 and residual plot) showed excellent fit of this model to the training set data. In Figure 6 all the data (training set, validation set and data predicted by other methods) that are available in the DIPPR database are plotted together with the values predicted by the linear QSPR. The predicted ΔH_f^0 value of methane (the first member of the series) is an outlier (15% prediction error, $n_C \geq n_{C\min}$), and is removed from further consideration. The average difference between the DIPPR data and the predicted values is 0.15%, and the maximal difference is 1.14% (for *n*-dotriacontane). Both differences are considerably smaller than the respective DIPPR data uncertainties. Thus, all the indicators show that this linear QSPR can confidently be used for interpolation and for both short and long-range extrapolation.

To predict ΔH_f^0 values for members of the target series the QSPR and the QPPR should be combined to yield: $\Delta H_f^0 = B_0 - 42.52 - 20.7\, n_C$ (kJ/mol) for $n_C \geq n_{C\min}$. For the series listed in Table 5 the B_0 and $n_{C\min}$ values provided in this table can be used. For other targeted series, $n_{C\min}$ should be determined, and the available ΔH_f^0 data for $n_C > n_{C\min}$ and the corresponding *n*-alkane data provided in Table 1 can then be used to determine the value of B_0.

Table 6. List of Compounds for which the DIPPR Recommended ΔH_f^0 Value was Replaced

No.	Series	Compound	n_C	Source of the data
1	1-alkene	propylene	3	Luria and Benson[15]
2		1-octene	8	Sakano et al.[14]
3	n-alkanoic acid	ethanoic acid	2	Pilcher and Cox[16]
4		propionic acid	3	Pilcher and Cox[16]
5	cis-2-alkene	cis-2-pentene	5	Westrum et al.[17]
6	trans-2-alkene	trans-2-pentene	5	Pedley et al.[18]
7		trans-2-nonene	9	Eliseev[19]

Prediction of ΔS_f^0 for the n-alkane series and the 12 target series

The ΔS_f^0 data for the *n*-alkane series (Table 7) can be used for derivation of a QSPR for the entropy of formation. Ten compounds in the range of $10 \leq n_C \leq 19$ are used as the training set. As in the case of the enthalpy of formation, ΔS_f^0 is also highly collinear with n_C. Fitting Eq. 1 (with $\zeta^* = n_C$ and $y = \Delta S_f^0$ kJ/(mol K), for $n_C > n_{C\min} = 3$), yielded the parameter values: $\beta_0 = (0.153541 \pm 2.5\text{e-}5)$, and $\beta_1 = (0.039217 \pm 1.96\text{e-}6)$ with a correlation coefficient $R^2 = 1$. All the statistical indicators (confidence interval on the parameters, R^2 and residual plot) showed excellent fit of this model to the training set as well as the evaluation set data. Expectedly, the highest prediction errors are obtained for methane (3.5%), and for ethane (1.24% since $n_C < n_{C\min}$), while the median prediction error for 31 members of the n-alkane series is 0.09% (when compared to the DIPPR experimental and predicted data).

Data related to ΔS_f^0 of the members of the 1-alkene series in the range of $1 \leq n_C \leq 20$ are shown in Table 7. To calculate the B_0 value for this property, the differences between the property values of the target and reference series $\delta y_{tr} = y_t - y_r = \Delta S_{ft}^0 - \Delta S_{fr}^0$ are calculated. The resultant δy_{tr} values (see Table 7) vary widely and even change sign. Most of the δy_{tr} values associated with experimental data are negative, except for 1-hexadecene, and most of the δy_{tr} values associated with predicted data are positive. For 1-hexadecene, replacing the DIPPR recommended value by another

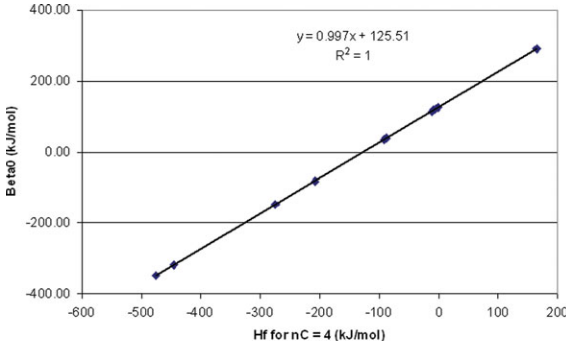

Figure 5. Plot of B_0 of various series vs. ΔH_{ft}^0 of the compound with $n_C = 4$.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

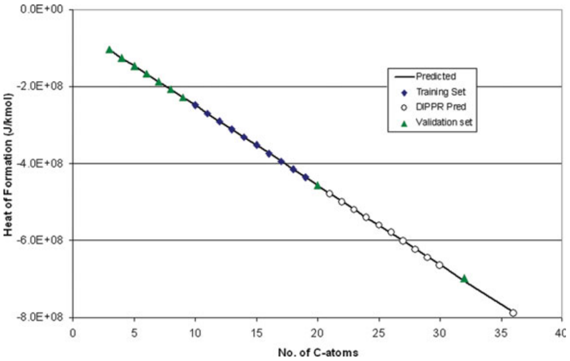

Figure 6. Prediction of ΔH_f^0 of *n*-alkanes with the linear QSPR: $\Delta H_f^0 = -42.52–20.7\ n_C$ (kJ/mol).

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

available experimental value (using data from Westrum et al.[15]) $\Delta S_f^0 = 0.7741$ kJ/(mol K), yields δy_{tr} = -0.0069 kJ/(mol K), where the sign is consistent with the rest of the experimental data. However, the differences between the δy_{tr} values still remain excessively large. The relative δy_{tr} values ($\delta y_{tr}/\Delta S_f^0$, shown in Table 7) are much smaller than the data uncertainty for the great majority of the 1-alkenes (except for propylene). This implies that the noise in the data is larger than the ΔS_f^0 difference between the target and reference series. Thus, using a QPPR even with $B_0 = 0$ (and $B_1 = 1$) yields prediction errors within the uncertainty limits.

Table 8 presents a summary of the ΔS_f^0-QPPR derived for the 12 homologous series. Note that for some of the series the minimal n_C considered for evaluating the parameter B_0 of the QPPR are different from the $n_{C\min}$ of respective values in Table 5 (because of differences in the availability of data and/or effect of the functional end group R). The range of the entropy of formation of the compounds from the various series at the minimal n_C is $0.2192 \leq \Delta S_f^0 \leq 0.3463$ kJ/(mol K). The range of the entropy of formation for the highest n_C compounds for which data are available is $0.5263 \leq \Delta H_f^0 \leq 0.9927$ kJ/(mol K).

The number of experimental/smoothed data points used for determination of B_0 varied between 15 (for *n*-alcohols) to 1 (*n*-alkanoic acids) or 2 (for cis-2-alkenes). The scarcity of experimental data for some of the series demonstrates the importance of the use of the ‘‘reference series’’ technique. The DIPPR recommended ΔS_f^0 values have proven to be consistent in all the tested cases and there was no need to replace any of the recommended values.

The range of the QPPR B_0 values is $-0.11 < B_0 < 0.067$ kJ/(mol K), with standard deviations (σ) in the range of $0.00016 \leq \sigma < 0.0012$ kJ/(mol K) (Table 8). Both the median and the maximal values of the deviations (%) between the QPPR predictions and the DIPPR experimental/smoothed data are < 1% for all 12 homologous series. The absence of excessively large deviations (that were present in the ΔH_f^0 predictions) can possibly be attributed to the same order of magnitude of the ΔS_f^0 data, where the absolute deviation results also in a similar order of magnitude for the relative deviation.

Based on the results presented in this section prediction of ΔS_f^0 for the (target) homologous series can be carried out using the equation $\Delta S_f^0 = B_0 + 0.153541 + 0.039217\ n_C$ (kJ/mol K) for $n_C \geq n_{C\min}$ with the parameters B_0 and $n_{C\min}$ shown in Table 8 (or determined by using experimental data of a target series).

Extension of the Reference Series Method to Model Additional Properties and Families of Compounds

The properties ΔH_f^0, ΔG_f^0, and ΔS_f^0 are defined at the constant temperature of 298.15 K, 1 bar and at a gaseous state. Therefore, in the case of homologous series, the values of these type II properties are expected to be collinear with the number of methylene groups (n) over a wide range of carbon numbers. The slope of the QPPR for these properties (for carbon number exceeding $n_{C\min}$.) is 1, based on physical considerations ($B_1 = 1$ in Eq. 2). This enables obtaining linear QSPRs and simple QPPRs, which accurately represent these properties in homologous series for any carbon number exceeding $n_{C\min}$.

For other type II properties (like heat of combustion, liquid molar volume, standard state enthalpy of formation, etc), the property data do not correspond to the same state for all members of the series for which data is available, due to

Table 7. Data and Relationship Between ΔS_f^0 of the *n*-alkane and 1-alkene Series

	ΔS_f^0 for n-alkanes[1]		ΔS_f^0.[1] for 1-alkenes[1]		$\delta y_{tr} = \Delta S_{ft}^0 - \Delta S_{fr}^0$	
No. of C-atoms	Value (kJ/mol K)	Uncertainty	Value (kJ/mol K)	Uncertainty	kJ/(mol K)	Relative (%)
1	**0.1863**	< 1%				
2	**0.2291**	< 1%	**0.2192**	< 3%	-9.92E-03	-4.53
3	**0.2702**	< 1%	**0.2670**	< 1%	-3.20E-03	-1.20
4	**0.3099**	< 1%	**0.3074**	< 3%	-2.51E-03	-0.82
5	**0.3495**	< 1%	**0.3462**	< 3%	-3.25E-03	-0.94
6	**0.3887**	< 1%	**0.3863**	< 3%	-2.44E-03	-0.63
7	**0.4280**	< 1%	**0.4252**	< 3%	-2.78E-03	-0.65
8	**0.4672**	< 1%	***0.4637***	< 3%	-3.53E-03	-0.76
9	**0.5064**	< 1%	0.5041	< 3%	-2.30E-03	-0.46
10	**0.5457**	< 1%	***0.5433***	< 3%	-2.40E-03	-0.44
11	**0.5849**	< 1%	**0.5822**	< 3%	-2.75E-03	-0.47
12	**0.6242**	< 1%	**0.6185**	< 1%	-5.65E-03	-0.91
13	**0.6634**	< 1%	0.6574	< 3%	-5.97E-03	-0.91
14	**0.7026**	< 1%	0.6966	< 3%	-5.99E-03	-0.86
15	**0.7418**	< 1%	0.7453	< 3%	3.49E-03	0.47
16	**0.7810**	< 1%	**0.7819**	< 1%	8.80E-04	0.11
17	**0.8202**	< 1%	0.8220	< 3%	1.77E-03	0.22
18	**0.8595**	< 1%	0.8594	< 3%	-5.00E-05	-0.01
19	**0.8987**	< 1%	0.9023	< 3%	3.64E-03	0.40
20	**0.9379**	< 1%	0.9386	< 3%	-2.98E-03	-0.32

Table 8. Summary of QPPR Derivation for ΔS_f^0 of 12 Series

			Range of ΔS_f^0 (used) kJ/mol		Data points[1]		QPPR		% Difference[2]	
No.	Series	n_{Cmin}	at n_{Cmin}	at (highest n_C)	Exp./Smoothed	predicted	B_0	σ	Median	Maximal
1	1-alkene	3	0.2192	0.9386(20)	10	8	−0.0031	9.92E-04	0.25	0.5
2	*n*-alcohol	2	0.28064	0.9894	15	5	0.0515	2.00E-04	0.01	0.29
3	*n*-mercaptan	2	0.2961	0.6901(12)	8	4	0.0661	6.00E-04	0.026	0.314
4	*n*-alkanoic acid	2	0.2825	0.9927(20)	1	18	0.0534	-	-	.-
5	*n*-acetate	4(5)	0.4023	0.6776(12)	3	5	0.0534	5.00E-04	0.06	0.14
6	1-alkyne	4	0.2904	0.5263(10)	3	4	−0.0195	1.60E-04	0.019	0.044
7	aldehyde	2	0.2642	0.6969(13)	11	2	0.0336	2.70E-04	0.029	0.58
8	*n*-formates	3	0.3282	0.6432(11)	3	6	0.0576	1.00E-03	0.141	0.225
9	*n*-amines	2	0.2848	0.7576(14)	12	-	0.0546	8.00E-04	0.06	0.447
10	cis-2-alkenes	(4)5	0.3463	0.6199(12)	2	5	−0.0027	6.00E-04	0.26	0.54
11	trans-2-alkenes	(4)5	0.3431	0.9281(20)	3	6	−0.0077	1.20E-03	0.38	0.58
12	alkyl-benzenes	7	0.321	0.939(24)	4	14	−0.1065	3.61E-04	0.117	0.187

[1] No. of compounds for which DIPPR recommended property values are available

[2] Differences between values predicted by the QPPR and the DIPPR recommended values. Respective heat of formation values are shown within brackets.

[3] The number in the brackets indicates the lowest n_C used for QPPR derivation when it differs from n_{Cmin}

phase change. Therefore, the same linear QSPR or QPPR cannot correctly represent the variation of these properties over the different states. Several other type II properties are subject to temperature and pressure change, which vary nonlinearly with n (for example $V_C = V_C$ (n, T_C, P_C), where T_C, P_C are type I properties). Such properties can be represented by linear QPPRs and QSPRs in terms of n (or n_C) only locally, and the slope of the QPPR is not necessarily 1. For long range extrapolation, descriptors that linearize the property variation in the series and/or nonlinear QSPRs and QPPR's have to be derived. Similarly, for type I properties that converge to a finite y^{∞} value, a constant incremental change can also be assumed only locally. For such properties, descriptors that linearize the property and/or nonlinear QSPRs and QPPR's have to be used for long-range extrapolation.

As noted by Peterson[8], properties of homogeneous series which are predicted by the linear group contribution methods correspond to linear QPPRs with a unit slope. The one to one correspondence between QPPRs with slope of 1, and the linear group contribution method enables critical assessment of the later, based on the available experimental data and theoretical considerations.

Peterson[7,8] recommended the use of the reference (or analogous) series method to a wider range of families of materials, such as families of organic molecules of the form R_1XR_2, containing different functional groups X with R_1 and R_2 representing either branched or linear alkyl groups. He[8] also demonstrated the use of the method for several properties not mentioned here: gas solubility in liquids, GC retention times, Henry's law constants and infinite activity coefficients. For these properties the *n*-alkane series cannot be used as reference series, and other reference/target pairs were used. Similarly to the conclusions reached in this study, Peterson[7,8] found that that the reference series method is particularly useful when one series has more or better data than do other series, and it is most beneficial for determining outliers and incorrect data (consistency analysis).

Conclusions

A two-stage procedure for analyzing property data and predicting properties of long-chain substances have been developed. The use of this procedure has been demonstrated for properties which are correlated with the number of methylene groups in homologous series: ΔH_f^0 and ΔS_f^0. The *n*-alkane series, for which the largest amount and highest precision data are available, was used as reference series and 12 series containing various functional groups were used as target series. The data available for the *n*-alkane series enables deriving a reliable linear QSPR to allow a long-range extrapolation to predict the properties of long-chain alkanes.

For the target series, consistency analysis of the available experimental data was carried out and data points were selected for removal or inclusion based on the consistency with the properties of the other compounds in the same series and in the reference series. As the slope of the property-property relation (QPPR) is set to 1 based on physical considerations, some reliable property data of the target series (corresponding to $n_C \geq n_{C\min}$) are needed only for determining the free parameter B_0 of the linear QPPR. To this aim, even a single reliable property value may be sufficient. For property values that change by several orders of magnitude within the range of interest, some excessive relative prediction errors were noticed in the low-absolute value region (mostly in low n_C region). The causes of these excessive errors still need to be investigated.

It was shown that using the proposed technique combined with a proper consistency analysis of the available experimental data, accurate and reliable QSPRs for predicting ΔH_f^0, ΔS_f^0 and ΔG_f^0 of long-chain substances can be derived. Some guidelines are provided for possible extension of the proposed method for properties which are collinear with the number of carbon atoms only locally.

Literature Cited

1. Marano JJ, Holder GD. General equations for correlating the thermophysical properties of *n*-paraffins, *n*-olefins and other homologous series. 1. Formalism for developing asymptotic behavior correlations. *Ind Eng Chem Res*. 1997;36:1895–1907.
2. Nikitin ED, Popov AP, Bogatishcheva NS. Critical properties of long-chain substances from the hypothesis of functional self-similarity. *Fluid Phase Equilibria*. 2005;235:1–6.
3. Paster I, Shacham M, Brauner N. Adjustable QSPRs for prediction of properties of long-chain substances. *AIChE J*. 2011;57(2):423–433.
4. Paster I, Brauner N, Shacham M. A "reference series" method for prediction of properties of long-chain substances. Paper 720h presented at:11th AIChE Annual Meeting, Oct. 16–21, 2011; Minneapolis, MN.
5. Paster I, Shacham M, Brauner N. Predicting a wide variety of constant pure compound properties for long chain substances using a "reference series" method. In I. D. L. Bogle, M. Fairweather, eds.

Proceedings of the 22nd european symposium on computer aided process engineering; June 17–20 2012; London, UK.
6. Marano JJ, Holder GD. General equation for correlating the thermophysical properties of n-paraffins, n-olefins and other homologous series. 2. Asymptotic behavior correlations for PVT properties. *Ind Eng Chem Res*. 1997;36:1887–1894.
7. Peterson BK. Relationships between the properties of families of materials. *Ind Eng Chem Res*. 2010;49(7):3492–3495.
8. Peterson BK. Property estimation using analogous series. *Ind Eng Chem Res*. 2011;50(12):7696–7704.
9. Rowley RL, Wilding WV, Oscarson JL, Yang Y, Zundel NA. *DIPPR Data Compilation of Pure Chemical Properties Design Institute for Physical Properties*. Provo, UT: Brigham Young University; 2010. http://www.aiche.org/dippr/.
10. Sellers P, Stridth G, Sunner S. An examination of the constancy of the CH_2 increment in homologous series. *J Chem Eng*. 1978;23:250–258.
11. Todeschini R, Consonni V. *Handbook of Molecular Descriptors*. Weinheim: Wiley-VCH; 2000.
12. Rowley RL. Personal communications; 2010.
13. Greenshields JB, Rossini FD. Molecular structure and properties of related compounds. *J Phys Chem*. 1958;62:271–280.
14. Sakano TK, Riddick JA, Bunger WB. *Organic Solvents: Physical Properties and Methods of Purification*. 4th ed. New York: Wiley Interscience; 1986.
15. Luria M, Benson SW. Electrostatics and the chemical bond. II. unsaturated hydrocarbons. *J Am Chem Soc*. 1975;97:3336–3342.
16. Pilcher G, Cox JD. *Thermochemistry of Organic and Organometallic Compounds*. New York: Academic Press; 1970.
17. Westrum EF Jr, Stull DR, Sinke GC. *The Chemical Thermodynamics of Organic Compounds*. New York: John Wiley and Sons; 1969.
18. Pedley JB, Naylor RD, Kirby SP. *Thermochemical Data of Organic Compounds*. London, UK: Chapman and Hall; 1986.
19. Eliseev NA. Thermodynamic study of dehydrogenation of C6–9 n-alkanes into alkenes. *Neftekhimiya*. 1985;25:755–760.

Manuscript received Jan. 5, 2012, and revision received Mar. 27, 2012.