engineering proceedings

Proceeding Paper

Group Contribution Revisited: The Enthalpy of Formation of Organic Compounds with "Chemical Accuracy" †

Robert J. Meier

Pro-Deo Consultant, 52525 Heinsberg, North-Rhine Westphalia, Germany; r.meier@planet.nl

† Presented at the 2nd International Electronic Conference on Processes: Process Engineering—Current State and Future Trends (ECP 2023), 17–31 May 2023; Available online: https://ecp2023.sciforum.net/.

Abstract: we established a Group Contributiuon (GC) parametrization for the heat of formation of organic molecules but, and this is new, revealing chemical accuracy (1 kcal/mol). Compared to previous approaches which did not achieve this, we succeeded by (i) taking reliable and consistent experimental data, (ii) not relying on computer-assisted automated parameter estimation, (iii) taking into account the physico-chemistry known for years, i.e., only introducing additional parameters when we understand the physico-chemistry, and finally (iv) acknowledging that the linear additive GC method has its limits and cannot account properly for any molecule. Not only averaged absolute deviations but also individual results were almost without exception within chemical accuracy, except for some more heavily substituted molecules for which the Group Contribution approach breaks down.

Keywords: enthalpy of formation; thermodynamics; process design; physico-chemical property prediction; group-contribution method; chemical accuracy; quantum chemistry

Citation: Meier, R.J. Group Contribution Revisited: The Enthalpy of Formation of Organic Compounds with "Chemical Accuracy". *Eng. Proc.* **2023**, *5*, x. https://doi.org/10.3390/xxxxx

Published: 31 May 2023

Copyright: © 2023 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The Group Contribution (GC) approach is a meanwhile old [1] approach and was, in the course of time, developed for many molecular properties. The basic idea is illustrated in the Scheme below where the molecule is broken into building blocks, the Groups, and we also see the Groups that constitute 2-methyl-1-propanol. The GC approach inherently assumes that the Groups behave independently and the value of a molecular property is calculated from the sum of the individual properties from the

$$\Delta H_f \text{ (molecule)} = \Sigma\ H_f \text{ Group Contribution of constituting Groups}$$

Groups by

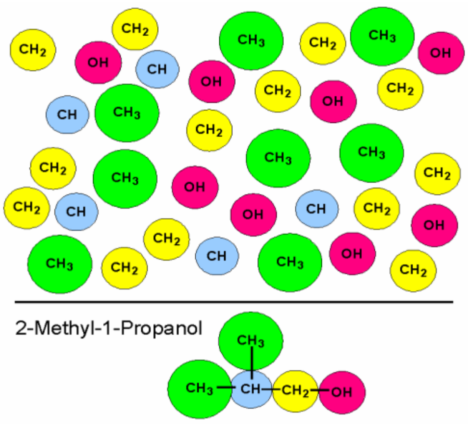

Figure 1. Caption.

These quick and easy to use GC methods, in particular when a Graphical Interface is provided (e.g., the ICAS program suite [2]), are very beneficial to chemical process developers. One of these properties is the heat of formation ΔH_f of organic molecules. To appropriately describe chemical transformations and equilibria, the heat of formation must have chemical accuracy, i.e., 1 kcal/mol or better. Moreover, a method must be reliable which means there should have no or really very few outliers. So not only the absolute averaged deviation of the model values should be small, but each individual value should be within chemical accuracy. Up till recently GC methods for the heat of formation of organic molecules did not achieve chemical accuracy [3,4].

2. Results

We started an attempt to revise the GC approach for ΔH_f and meanwhile we have succeeded to achieve this goal [5–8]. Compared to previous approaches this positive outcome was the result of a number of specific actions. First of all, in the procedure to determine the numerical values for the Group Contribution parameters we almost exclusively used reliable and consistent experimental data selected after consulting experts on experimental thermodynamics, which is crucial because of the 1 kcal/mol requirement. Previous attempts have often relied on the use of experimental heats of formation from data bases, but these are generally collections of data from a large variety of literature sources without critical screening. Recently, Chan reported [9] that for the NIST data base [10] the heat of formation of only less than 40% of all species included were found within chemical accuracy. Moreover, even a pretty normal organic molecules such as 2-methyl-4-methylene-1,3-dioxolane and 2,4-dimethyl-1,3-dioxole were found off by more than 100 kJ/mol. When we use reliable data from proper sources, i.e., those from Rossini and co-workers the linear alkanes [11], we observe a very constant increment and a very small averaged absolute difference (AAD) between experiment and GC model: 0.19 kJ/mol. It is interesting to note that when we take further experimental data from other sources for longer alkanes, i..e. heneicosane up till and including hexatriacontane, the AAD for these only is over 1 kJ/mol which, though as such still very good, is much larger than for the consistent Rossini data and that for simple species such as the linear alkanes. Next to the GC parameter value for the CH_2 Group of −20.63 kJ/mol, at the same time we have obtained a unique GC parameter value for the CH_3 Group: −42.36 kJ/mol.

Secondly, in order to achieve chemical accuracy, we observed that we needed to determine the parameters step-by-step, in fact by hand, so we could identify specific deviations. Most of the time researchers use automated procedures in which larger amount of data are optimized simultaneously, but this makes it virtually impossible to trace and understand specific deviations. One obtains a mathematical fit without the guarantee that the correct chemistry is represented. Rather than looking at individual data, more parameters are introduced to obtain better agreement with experimental data, which easily leads to overfitting and incorrect predictions for molecules which were not involved in establishing the Group Contribution parameter values. Our procedure, as we learned it through our recent investigations [5–8], resulted in an absolute minimum number of GC parameter, thereby avoiding overfitting, as additional parameters are only added based on concrete information on deviations between model and experimental data and understood based on physico-chemical information. We have illustrated this with various examples [5–8].

Starting from the linear alkanes, the next step were the methyl branched alkanes. Analysis of the experimental data in conjunction with the GC model revealed we can achieve good model results when adding nearest neighbour interactions. Using the concept of nearest and next-nearest neighbour interactions is quite common in physics and chemistry and these can, if desired, also be confirmed by ab initio quantum calculations [8]. The latter makes we have a good physico-chemical justification for these additional parameters. Various neighbour interaction parameters we thus established are shown in the Scheme below, where the one in yellow is the 1,5-Me-Me-interaction.

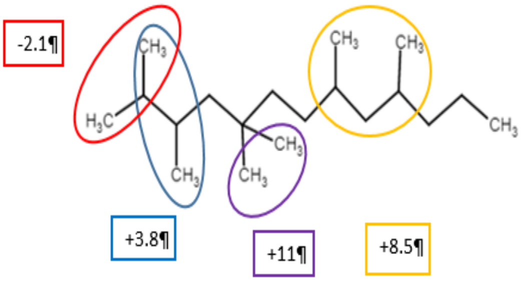

Figure 2. Caption.

An interesting example is formed by the nitriles. Whereas for the mononitriles we see good agreement between the model and experiment, for the dinitriles this is not the case for the first two in the series, viz. Table 2. Various models describe these two well but require additional (not a priori justified) parameters to account for the higher dinitriles. In reality one should look at it exactly the other way around: malono- and butanedinitrile are distinct cases because they are subject to so-called germinal effects. These were described in detail by Beckhaus et al. [12] who reported a synergetic destabilisation by two relatively close geminal cyano substituents of 48 kJ/mol, which is relatively close to the deviation of almost 55 kJ/mol presented in Table 1. So also here, accounting for the correct physico-chemistry leads to a good and scientifically justified model.

Table 1. Experimental [10] and model values for mono- and dinitriles. All values in kJ/mol. The AAD for the dinitriles is based on the two last species only, for discussion see text.

Nitriles	NIST	Model dHf	Model-Exp	ABS (Model-Exp)
propanenitrile	51.5	53.01	1.51	1.51
butanenitrile	31.2	32.38	1.18	1.18
pentanenitrile	11.1	11.75	0.65	0.65
hexanedinitrile		−29.51	1.45	1.45
octanenitrile	−50.6	−50.14	0.46	0.46
decanenitrile	−91.6	−91.4	0.20	0.2
tetradecanenitrile	−174.8	−173.92	0.88	0.88
averaged absolute difference				0.80
malononitrile	266.3	211.37	−54.93	54.93
butanedinitrile	209.7	190.74	−18.96	18.96
pentanedinitrile		170.11	0.11	0.11
hexanedinitrile	149	149.48	0.48	0.48
averaged absolute difference				0.30

Another case for which it was not possible to get good agreement between model and experimental values, within chemical accuracy, were alkyl-ethers. However, when we calculated the geometries using quantum calculations (B3LYP) we could establish a relation between the COC valence angle and the magnitude of the GC parameter value associated with the ether Group. The relevant data have been collected in Table 2 and we observe very good results within chemical accuracy.

Table 2. Experimental and model values for various ethers. All values in kJ/mol.

Methyl-alkyl-ethers	Verevkin [13]	Model dHf	Model-Exp	ABS (Model-Exp)	Ether Group Constitution	GC Value ether Group	COC Valence Angle
dimethylether	−184.1	−184.1	0.00	0.00	Me-O-Me	−184.1	112.7
methyl ethyl ether	−216.4	−217.36	−0.96	0.96	Me-O-C-R	−175	113.1
methyl propyl ether	−238.4	−237.99	0.41	0.41	Me-O-C-R		113.1

methyl n-butyl ether	−258.3	−258.62	−0.32	0.32	Me-O-C-R		113.1
methyl decyl ether	−381.1	−382.4	−1.30	1.30	Me-O-C-R		113.1
methyl isopropyl ether	−252	−252.72	−0.72	0.72	Me-O-CRR′	−168	115.1
methyl t-butylether	−283.4	−282.08	1.32	1.32	Me-O-CRR′R″	−156	118.4
methyl t-amylether	−301.1	−302.71	−1.61	1.61	Me-O-CRR′R″		118.7
averaged absolute difference				0.83			
Di-alkyl ethers	**Verevkin 2002**	**Model dHf**	**Model-Exp**	**ABS (Model-Exp)**	**Ether Group Constitution**	**GC Value ether Group**	**COC Angle**
diethylether	−252.1	−252.72	−0.62	0.62	R-COC-R′	−168	113.5
ethyl propyl ether	−272.4	−273.35	−0.95	0.95	R-COC-R′		113.5
ethyl butyl ether		−293.98			R-COC-R′		113.5
di-n-propylether	−293.1	−293.98	−0.88	0.88	R-COC-R′		113.5
di-n-butylether	−332.9	−335.24	−2.34	2.34	R-COC-R′		113.7
di-n-pentylether	−380.4	−376.5	3.90	3.90	R-COC-R′		113.4
ethyl t-amylether	−333.5	−336.07	−2.57	2.57	R-COC-R′R″R	−146	119.4
butyl t-amylether	−375.7	−377.33	−1.63	1.63	R-COC-R′R″R‴		119.1
ethyl t-butylether	−316.8	−315.44	1.36	1.36	R-COC-R′R″R‴		118.8
propyl t-butylether	−339.3	−336.07	3.23	3.23	R-COC-R′R″R‴		118.7
n-butyl t-butylether	−360.1	−356.7	3.40	3.40	R-COC-R′R″R‴		118.6
amyl t-butylether	−380.6	−377.33	3.27	3.27	R-COC-R′R″R‴		
di-i-propylether	−319.4	−318.44	0.96	0.96	RR′-COC-R″R‴	−149	116
di-sec-butylether	−361.3	−359.7	1.60	1.60	RR′-COC-R″R‴		116.5
t-butyl s-butylether	−379	−381.43	−2.43	2.43	RR′-COC-R″R‴R″″	−149	119.8
t-butyl i-propylether	−360.1	−360.8	−0.70	0.70	RR′-COC-R″R‴R″″		119.9
t-butyl i-butylether	−367.9	−364.8	3.10	3.10	RR′-COC-R″R‴R″″		119
di-t-butylether	−361.2	−361.16	0.04	0.04	tBU-COC-tBu	−107	128
averaged absolute difference				1.94			

A further crucial aspect is the relevant size of the chemical Groups rather than using the smallest possible entities which is often done, e.g., an aromatic carbon atom in a benzene ring rather than the benzene ring as Group. In various cases the definition of Groups which are not the smallest enabled very good results, i.e., chemical accuracy, which could not be obtained otherwise. It should not be surprising that we need to define larger entities as Groups for, as a typical example, (heterogeneous) aromatic ring systems. The electronic interactions make that the individual atoms (e.g., the CH entity cannot be regarded as an independent entity), and thus the GC concept breaks down at least when we want to achieve good accuracy of prediction. Thus, in the case of pyridines and quinolones we adopted these two molecules as Groups themselves. When we combined this with our already established GC model (parameters for other Groups) we obtained very good agreement between model and experiment, viz. Table 3.

Table 3. Experimental [14] and GC model heats of formation for selected pyridines and quinolones. All values in kJ/mol. Both pyridine and quinoline have been adopted as a Group themselves and are therefore indicated with an *.

Pyridines and Quinolines	Verevkin [14]	Model dHf	Model-Exp	ABS (Model-Exp)
pyridine	140.4	142 *	1.60	1.60
2-methylpyridine	99.2	99.64	0.44	0.44
3-methylpyridine	106.4	105.64	−0.76	0.76
4-methylpyridine	104.1	105.64	1.54	1.54
2,3-dimethylpyridine	68.3	63.28	−5.02	5.02
2,4-dimethylpyridine	63.9	63.28	−0.62	0.62
2,5-dimethylpyridine	66.5	63.28	−3.22	3.22
2,6-dimethylpyridine	58.7	57.28	−1.42	1.42
3,4-dimethylpyridine	70.7	75.78	5.08	5.08

3,5-dimethylpyridine	72.8	75.78	2.98	2.98
2-ethylpyridine	75.6	79.01	3.41	3.41
3-ethylpyridine	82.9	85.01	2.11	2.11
4-ethylpyridine	80.6	85.01	4.41	4.41
quinoline	200.5	197 *	−3.50	3.50
2-methylquinoline	156.6	160.64	4.04	4.04
4-methylquinoline	158.6	160.64	2.04	2.04
6-methylquinoline	157.3	160.64	3.34	3.34
8-methylquinoline	164.8	160.64	−4.16	4.16
2,6-dimethylquinoline	121.3	124.28	2.98	2.98
2,7-methylquinoline	119.8	124.28	4.48	4.48
2-phenylquinoline	286.6	287.5	0.90	0.90
averaged absolute difference				2.76

Finally, it is important to realize and to make explicit that certain effects, i.e., certain molecules, cannot be treated by a simple linear additive method which the GC method is. One example are the (substituted) cycloalkanes, where we have shown that using quantum calculations of the G4 type [15] we can establish trends and we can extract additional parameters to be added to the GC model leading to model values within chemical accuracy [8]. Another example are heavily substituted alkyl-substituted species. Whereas 3,5,5-trimethyl-1-hexanol and 2,2,4,4-tetramethyl-3-pentanol were well-accounted for, this was not the case for 2,2,4,4-tetramethyl-3-iPr-3-pentanol (33 kJ/mol deviation). We could show that using B3LYP quantum calculations one can find out whether a case cannot be treated with the linear additive GC approach [8].

3. Conclusions

In summary, we established a new GC parametrization for the heat of formation of organic molecules for which not only averaged absolute deviations but also individual results were almost without exception within chemical accuracy, except for some more heavily alkyl-substituted molecules for which the Group Contribution approach breaks down.

Acknowledgments: The author sincerely thanks Sergey Verekvin (University of Rostock) and Jurgen Rarey for discussion and advice on the quality of experimental data, Paul Rablen (Swarthmore College, Swarthmore, PA, USA) for the joint work reported in Reference [8], and Georgios Kontogeorgis and Gürkan Sin (Danisch Technical University, DTU) for allowing the use of the ICAS23 software suit.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Van Krevelen, D.W.; Chermin, H.A.G. Estimation of the free enthalpy (Gibbs free energy) of formation of organic compounds from group contributions. *Chem. Eng. Sci.* **1951**, *1*, 66–80; Erratum in *Chem. Eng. Sci.* **1952**, *1*, 238.
2. ICAS23 Software Package Which Includes the Marrero-Gani Method as Well as the Constantinou-Gani Approach and Also Provides Results Based on the Joback and Reid GC Method. Available online: https://www.kt.dtu.dk/english/research/kt-consortium/software (accessed on 21 February 2023).
3. Hukkerikar, A.S.; Meier, R.J.; Sin, G.; Gani, R. A method to estimate the enthalpy of formation of organic compounds with chemical accuracy. *Fluid Phase Equilibria* **2013**, *348*, 23–32. https://doi.org/10.1016/j.fluid.2013.03.018.
4. Kadda, A.; Mustapha, B.A.; Yahiaoui, A.; Khaled, T.; Hadji, D. Enthalpy of Formation Modeling Using Third Order Group Contribution Technics and Calculation by DFT Method. *Int. J. Thermodyn.* **2020**, *23*, 34–41. https://doi.org/10.5541/ijot.647800.
5. Meier, R.J. Group contribution revisited: The enthalpy of formation of organic compounds with “chemical accuracy”. *ChemEngineering* **2021**, *5*, 24. https://doi.org/10.3390/chemengineering5020024.
6. Meier, R.J. Group contribution revisited: The enthalpy of formation of organic compounds with “chemical accuracy” Part II. *AppliedChem* **2021**, *1*, 111–129. https://doi.org/10.3390/appliedchem1020009.
7. Meier, R.J. Group contribution revisited: The enthalpy of formation of organic compounds with “chemical accuracy” Part III. *AppliedChem* **2022**, *2*, 213–228. https://doi.org/10.3390/appliedchem2040015.

8. Meier, R.J.; Rablen, P.R. Group contribution revisited: The enthalpy of formation of organic compounds with "chemical accuracy" Part IV. *Thermo* **2023**, *submitted*.
9. Chan, B. High-Level Quantum Chemistry Reference Heats of Formation for a Large Set of C, H, N, and O Species in the NIST Chemistry Webbook and the Identification and Validation of Reliable Protocols for Their Rapid Computation. *Phys. Chem. A* **2022**, *126*, 4981−4990. https://doi.org/10.1021/acs.jpca.2c03846.
10. NIST Data Base. Available online: https://webbook.nist.gov/ (accessed on 16 March 2023).
11. Prosen, E.J.; Rossini, F.D. Heats of combustion and formation of the paraffin hydrocarbons at 25 °C. *J. Res. Natl. Bur. Stand.* **1945**, *34*, 263–269.
12. Beckhaus, H.-D.; Dogan, B.; Pakusch, J.; Verevkin, S.; Rüchardt, C. Abhangigkeit des inversen anomeren Effektes geminaler Nitril-Gruppen von der Struktur, *Chem. Ber.* **1990**, *123*, 2153–2159.
13. Verevkin, S.P. Improved Benson Increments for the Estimation of Standard Enthalpies of Formation and Enthalpies of Vaporization of Alkyl Ethers, Acetals, Ketals, and Ortho Esters. *J. Chem. Eng. Data* **2002**, *47*, 1071–1097. https://doi.org/10.1021/je020023o.
14. Verevkin, S.P.; Safronov, S.P.; Samarov, A.A.; Vostrikov, S.V. Hydrogen Storage: Thermodynamic Analysis of Alkyl-Quinolines and Alkyl-Pyridines as Potential Liquid Organic Hydrogen Carriers (LOHC). *Appl. Sci.* **2021**, *11*, 11758. https://doi.org/10.3390/app112411758.
15. Rablen, P. A Procedure for Computing Hydrocarbon Strain Energies Using Computational Group Equivalents, with Application to 66 Molecules. *Chemistry* **2020**, *2*, 22. https://doi.org/10.3390/chemistry2020022.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.