Manual associated with an Excel working document ('GC model') and belonging to the Group Contribution approach for the heat of formation of organics

The current Group Contribution (GC) method was developed with the intention to have a GC method with chemical accuracy, i.e. the difference between experimental and model value less than 1 kcal/mol or 4.2 kJ/mol. This was unprecedented (in 2024) and matches the needs for proper and reliable process design. Moreover, attention was paid that not only the root-mean-square deviation is within chemical accuracy, but each individual model heat of formation should be within chemical accuracy from its experimental value. The heat of formation refers to the gaseous state and 298.15K. Thus far, see the publications below, chemical accuracy was achieved except for specific cases which have been addressed specifically. The reasons for failure in specific cases are given, e.g. steric hindrance, and alternatives (other than the GC method) how to obtain chemical accuracy provided.

The scientific publications which describe the current GC method include

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
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
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
Meier, R.J.; Rablen, P.R. Group contribution revisited: The enthalpy of formation of organic compounds with "chemical accuracy" Part IV. *Thermo* **2023**, *3*, 289–308. https://doi.org/10.3390/thermo3020018
Meier, R.J.; Rablen, P.R. Group contribution revisited: The enthalpy of formation of organic compounds with "chemical accuracy" Part V. *Applied Sciences* **2024** 14, 1929. https://doi.org/10.3390/app14051929
Meier, R.J.; Rablen, P.R. Group contribution revisited: The enthalpy of formation of organic compounds with "chemical accuracy" Part VI. **2024** Current paper with this Manual as Supplementary Material.

The last publication comprises this Manual as well the Excel file with all data as Supplementary Material. ***Users are requested to refer to these works in any kind of publication of data generated using the GC method developed here.*** The latter also enable others to evaluate the heat of formation for another molecule by adding the appropriate data to the file. It is important to realize that chemical accuracy can be achieved when staying within the space of molecules which were considered during our investigations, viz. the papers cited above. When going beyond one should expect similar accuracies, beyond chemical accuracy, as is the case for other currently available GC models.

When while applying the GC method with the aid of the Excel file in one way or the other, in case specific questions come up please look up possible details in the relevant publication, the first column of the Excel file contains the number of the publication in which a specific class of molecules was treated.

Predicting the gas-phase heat of formation: how to proceed in practice

The generic form associated by a Group Contribution method (in its simplest form) reads

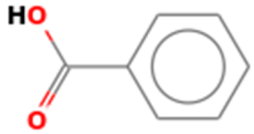
$$\Delta H_f = \sum_{j=1,N} Nj \cdot \Delta H_f(j) \qquad (1)$$

in which Nj represents the number of times Group j occurs in the molecule of interest, whereas $\Delta H_f(j)$ is the Group Contribution of the chemical Group j to the heat of formation. ΔH_f is the enthalpy of formation for a molecule, more precisely the ideal gas species at the reference temperature of 298.15 K. And so, as an example, for n-alkanes and

$$\Delta H_f \text{ (n-alkanes)} = 2 * \Delta H_f (CH3) + N_{CH2} * \Delta H_f (CH2) \qquad (2)$$

More generically (see the Excel file) it could be J11*J104+K11*K104+M11*M104, with J11 the GC parameter for CH3, K11 the parameter for CH2, and M11 for C (tertiary carbon). The variables J104, K104 and M104 are the number of times each of the three Groups occurs in the molecule.

In our method we have additional interaction energies, in particular (nearest) neighbour interaction energies. These neighbours can be bonded neighbours such as in



and the heat of formation is given by ΔH_f (phenyl) + ΔH_f (carboxylic) + phenyl-carboxylic acid interaction. This leads to the generic GC formulae for our parametrisation

$$\Delta H_f = \sum_{j=1,N} Nj \cdot \Delta H_f(j) + \sum_{j,k=1,N} \Delta H_f(j,k) \qquad (3)$$

Now for your molecule of interest, you may proceed in one of the following two, slightly different, ways.

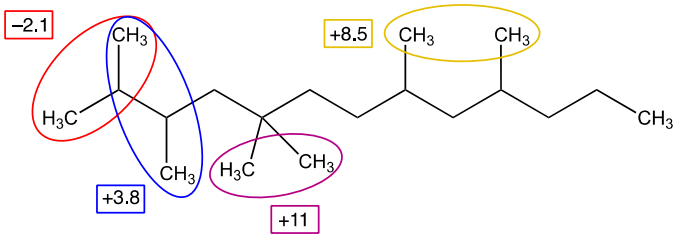
(1) Add the molecule at an appropriate place in the accompanying Excel table, at a place most close to a similar molecule. This may help in the selection of the correct GC parameters as in the column model dHf you can see the individual calculations and thus the GC contributions when gong to the appropriate cell in the table. It is particularly important to correctly account for specific Group-group interactions, as great care is required to include possibly required additional interaction energies the numerical values which are given at the top of the Table, lines 1-8 (and these are also mentioned in this Manual further below).

(2) Use the formula (3) and only looking at lines 1-13 in the Excel file comprising all GC parameters which have arisen from our studies.

Annotations on additional parameters accounting for Group-Group interactions

We will pay some extra attention to the Group-Group interaction parameters, so all that goes beyond Eq.(1) because the heat of formation cannot be appropriately accounted for by imply adding single isolated Group Contributions. Therefore rather than applying Eq.(1) we need to resort to Eq.(3). We will show a series of examples to illustrate.

The additional parameters are basically nearest or next-nearest neighbour interaction based. Examples include methyl-methyl interactions when attached to an alkyl:



For 2,3-dimethylhexane this leads to

ΔH_f (2,3-dimethylhexane) = 4 * GC_{CH3} + 2 * GC_{CH2} + 2 * GC_{CH} -2.1 + 3.8 kJ/mol

For the 2-alkenes with a substituent at the double bond, e.g. 2-methyl-2-butene, we need to add a correction of -11 kJ/mol to obtain the excellent results shown in Ref.1 Table S12. The central Group here is trans R-C=C-R'. A correction: whereas in Ref.1 (see top) the data in the Supplementary Materials Table S12 are the correct one, the formulae, nr. (11), in the text

ΔHf (2-alkenes with a substituent at the double bond) = NCH3 * GCCH3 + NCH * GCCH+ NCH2 * GCCH2 + GCC=C

is not quite correct as it contains $GC_{C=C}$ whereas it should have been $GC_{R\text{-}C=C\text{-}R'}$. Also when using NCH3 * GCCH3 + NCH * GCCH+ NCH2 * GCCH2 + GCC=C good results were obtained, but less good, and more importantly less consistent with the treatment of the other 2-alkanes.

For benzenes we have determined correction parameters for mono- (+6 kJ/mol), di- (+18.5 kJ/mol), tri-substituted (+30 kJ/mol) and tetra- (+40 kJ/mol) substituted alkylbenzenes. Thus, for example

ΔH_f (1,3-dimethylbenzene) = 2* GC_{CH3} + 1 * GC_{phenyl} +18.5 kJ/mol.

For the presence of an ethyl (similar for propyl) Group added onto an alkyl chain an interaction energy contribution of +6 kJ/mol needs to be added, so for 3-ethylhexane we have

ΔH_f (3-ethylhexane) = 3 * GC_{CH3} + 4 * GC_{CH2} + 1 * GC_{CH} + 6 kJ/mol

We simply do not have the information how this materializes with for instance a butyl side group, likely similar but no validated results yet.

When a benzene is alkyl substituted we need to add a correction depending on the degree of substitution. Correction factors alkyl substituted benzenes:

AromMonoalkyl	AromDialkyl	AromTrialkyl	AromTetraalkyl
+6	+18.5	+30	+40

The cycloalkanes and cycloalkenes suffer from ring strain which is a non-additive quantity.
The general formula to evaluate the GC heat of formation reads

ΔH_f (cycloalkane) = N_{CH2} * GC_{CH2} + ring strain

For cyclopropanes, also the methyl-substituted ones, a ring strain related energy correction of 115 kJ/mol is to be added.

For cyclobutanes, add a ring strain of 102.5 kJ/mol for the substituted cyclobutanes, and 110 kJ/mol for cyclobutane.

For the ring strain in cyclopentanes it is sufficient to adopt, as for the cyclopropanes, a single value for all cyclopentanes, for the cyclopentanes its value is 26.5 kJ/mol.

For the cyclohexanes a constant ring strain of -2 kJ/mol (note it is indeed a minus sign!) is to be added. Agreement within chemical accuracy between model and experiment for di-and tri-methylsubstituited cyclohexanes is obtained when adopting -3.6 kJ/mol for the ring strain.

For the chloromethanes one should adopt the experimental values. For the mono-chloro molecules GC parameters were given in the Table above. However, for 1,1-dichloro and 2,2-dichlroro we need to add found an additional correction of 16.6 kJ/mol and for 1,2-dichloro a correction (= nearest neighbour interaction energy parameter) of +6 kJ/mol. As an example, hexachloroethane has two CCl_3 groups, and within each of these groups we find three 1,1-interaction, and between the two CCl_3 groups we have 3 x 3 = 9 1,2-interactions. Herewith we can now calculated the heat of formation of hexachloroethane

ΔH_f (hexachloroethane) = 2 * GC_{C} + 6 * $GC_{Cl\ terminal}$ + 6 * 16.6 +9 * 6 (kJ/mol)

leading to a difference between model and experimental values of 3.2 kJ/mol only.

For fluoro hydrocarbons we find similar trends as for the chloroalkanes. For the mono-substituted it is all straightforward, but for e.g. 1,1,-difluoro alkanes an additional 1,1-interaction energy of -17.5 kJ/mol was established. For 1,1,1-trifluoro we have three times a 1,1-interaction and thus 3 x -17,.5 = -52.5 kJ/mol correction due to neighbour interactions, and this leads to very good agreement with the experimental heat of formation (difference 2.7 kJ/mol).

The current model does not (yet) provide acceptable results for aromatic fluoro species. Aromatic chloro compounds were not included yet at all, and therefore the current method should not be applied to such species.

Also for dinitro compounds with the nitro groups as neighbours (e.g. 1,1-, 1,2- and 2,2- dinitro) an additional interaction energy correction needs to be added. Due to a lack of sufficient data, the magnitude could, thus far, only be set for the 1,1-dinitro species.

When substituents other than an alkyl Group are attached to a benzene ring, we need to add a correction term accounting for the effects of the bond between the substituent and the benzene ring.

COOH to benzene ring correction +12 kJ/mol

t-BU to benzene ring +13.5 kJ/mol

amine-benzene -7 kJ/mol

OH-phenyl group correction +7 kJ/mol

Acetyl-benzene interaction +6 kJ/mol

1,2 methoxy interaction

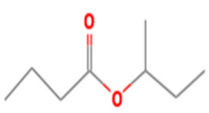
methoxy Groups nearest neighbour positions at benzene ring +11 kJ/mol

3,5 methoxy interaction

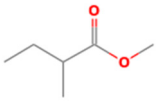
methoxy Groups at next-nearest neighbour positions at benzene ring -4.1 kJ/mol

1 methyl substitution at indole and indoline -7 kJ/mol correction

For aliphatic esters, whenever the branch of an alkyl is directly attached to the ether oxygen as in



contrary to a branch at the keto side as in



it is necessary to add a substituent correction factor of -9.5 kJ/mol to the initial GC result to obtain the final, correct GC model value.

Finally, for methylamines we observed a deviation between GC model values and experimental c.q. G4 values (based on results we obtained on species including methylamine (-6.4 kJ/mol), dimethylamine (-14.1 kJ/mol), trimethylamine (-10.9 kJ/mol), N-methyl-butanamine (-8.8 kJ/mol when compared to the G4 result) and *tert*-butylmethylamine (-14.7 kJ/mol). From a pragmatic point of view when adding +11 kJ/mol to all methyl substituted species we have agreement between experiment and GC values almost chemical accuracy for all five named species.

Highly important: when can the current model be applied

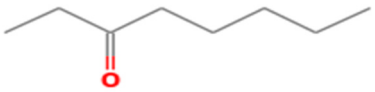
It is important to realize that, as we pursue chemical accuracy, we currently do not have GC parameters to allow the modelling of the heat of formation of any molecule within chemical accuracy. Up till now the work was successful in this respect and therefore there is no foreseen problem extending the method to many more classes of molecules. We recommend reading the appropriate publications (list of five in the above) depending on the type of molecule of interest. The molecule classes discussed in each of these publications are indicated in column 1 in the accompanying Excel spreadsheet 'GC model'. This is relevant as the text in the publications also points out when chemical accuracy cannot be achieved (with the current parametrisation).

In brief, a molecule which resembles a molecule in the Excel spreadsheet can normally be expected to be treated leading to a model value for ΔHf within chemical accuracy. We can thus treat alkanes

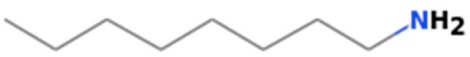




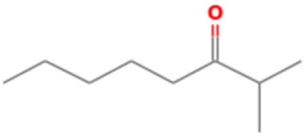
with confidence. We can also treat aldehydes and ketones



with confidence, as well as amines



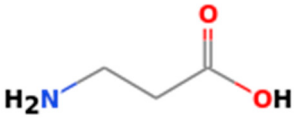
as long as substituents are not neighbours, so NOT as in



Because it could be that the keto and the methyl substituent require an additional interaction parameter. We can treat such a molecule with confidence when we do not have any near neighbour interactions. In effect this depends on each air of substituents. Potentially one would need a GC parameter for each Group enhanced by further interaction parameters for each potential pair.

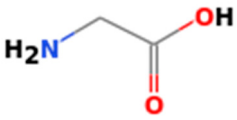
Below we present results, using our current GC approach, for a number of ***amino acid related species.***

For **β-alanine** we do not expect any additional issues as the functional groups are clearly separated.



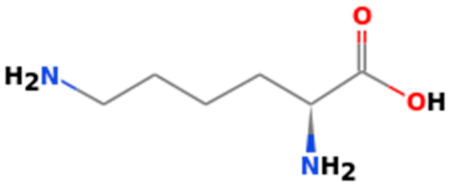
The experimental value of -424 ± 2 kJ/mol https://webbook.nist.gov/cgi/cbook.cgi?ID=C107959&Units=SI&Mask=1) is to be compared the GC model value which is established by one carboxylic Group, two methylene Groups and a terminal amino Group: -391 – 2*(20.63) +13 = -419.3 kJ/mol, and this is close to chemical accuracy considering experimental accuracy.

Glycine



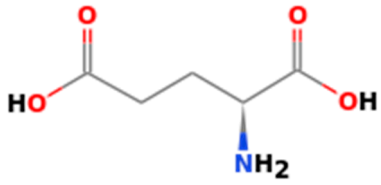
Here we may expect a potential interaction between the amino Group and (one of) the oxygens. For the GC heat of formation we have a carboxylic Group, one methylene Group and a terminal amino Group which adds up to -391 -20.63 +13 (these numerical values can be found in the Excel file) = -398.6 which differs by 6.5 kJ/mol from the experimental value -392.1 ± 0.6 given by Pedley c.s. 1986 (all references are quoted in the published papers).

Lysine (2,6-diaminohexanoic acid)



We start from aniline, -424.4 kJ/mol, and add one more amino Group and 4 more methylene Groups and subtract one methyl Group: -424.4 kJ/mol +13 -4*20.63 + 42.36= -451.6 kJ/ml. VanderSpoel c.s. have reported a G4 value of -448.1 kJ/mol, a difference of 3.5 kJ/mol.

Glutamic acid (amino pentanedioic acid)



This is aspartic acid with one more methylene Group, leading to a GC model value of -793.6 -20.63 = -814.2 kJ/mol, to be compared to a G4 value of -810.6 kJ/mol reported by vanderSpoel, a difference of 3.6 kJ/mol.

Another type of neighbour interaction to be accounted for is the bonding between two Groups, A-B. Not any combination leads to an additional nearest neighbour interaction energy parameter. But for instance an alkyl Group attached to a phenyl rings requires the Group Contributions for the alkyl and the phenyl Groups enhanced by an interaction parameter which, for a mono-alkyl substitution, amounts +6 kJ/mol. For di-alkyl substitution the additional interaction energy equals 18.5 kJ/mol. We

have created an additional Excel spreadsheet which reveals interaction energies to be taken into account between Groups as indicated like

	CH_3	Phenyl
CH_3		6
Phenyl	6	

In particular if larger substituents are neighbour substituents such as 1,2-di-substituted benzenes, e.g. 2-methylacetophenone. The deviation from experiment may amount for instance 10 or 12 kJ/mol, which is beyond chemical accuracy, but not necessarily worse than other existing GC methods. So the current method might be used as well, realizing however that very highly likely chemical accuracy is not achieved.

The publications give detail about cases which are problematic. These also include heavily congested branched alkanes such as 2,2,3,4,4-pentamethylpentane.

Alternatives when a molecule cannot be treated with current GC approach assuring chemical accuracy

When a molecule cannot be dealt with appropriately, mainly due to steric reasons but also other reasons might apply, you may choose one of the following options to move forward

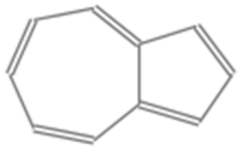
- Use a G4 quantum results, either from literature data or by performing a calculation, the later requires the tools and expertise however; the result is very likely within chemical accuracy
- When steric effects are expected to be involved, B3LYP calculations may be applied to calculate the relative energies between the molecule of interest and a second molecule in which the substituents are differently attached avoiding the steric overlap. Example: when 2-methylacetophenone is the target molecule, it can be compared (relative energy) to 3-methylacetophenone, as the latter is expected to be well-represented by the current GC model. The model value for 2-methylacetophenone is now given by the GC model value for 3-methylacetophenone + (B3LYP difference between 2-methylacetophenone and 3-methylacetophenone)
- Another option to apply B3LYP calculations, when G4 calculations are not within possibilities (molecule too large for that method or competence to perform such more expert type calculations not present), is to obtain an estimate for interaction parameters, e.g. as in diones. Ketones can be handled very well by our method, and so diones with the keto Groups far enough apart such as in 2,9-decanedione will be represented well. However, when we have the keto Groups more close such as in 4,5-decanedione or 3,5-decanedione there are additional interactions. Relative energy differences between 2,9-decanedione and the other configurations can normally be well evaluated by the B3LYP method (we suggest the larger 6-311+G** basis set).

Some explicit particular examples

Here we present some examples which are not quite standard when compared to the list in the Excel table, but these provide evidence that other molecules can be treated well with a

model heat of formation within chemical accuracy under the assumption that an appropriate base structure is used.

Example 1. Azulene



When we start from the recommended value for 1,3,5-cycloheptatriene +180.9 kJ/mol (Pedley et al.) we obtain

+180.9 + 108.9 (1,3-butadiene) + 5-membered ring strain + 2*conjugation = 289.8 + ring strain +22.6 - 2*17 (F942 conjugation)=278.4

When we start from 1,3,5-heptatriene, we add

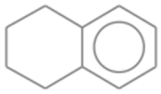
+180.9 + 133- 3*17 + 20.6 + 4= +287.5

which is very close to the experimental value of 289.1 kJ/mol given by Pedley et al.

Note: such an analysis should, for instance, not be attempted to obtain the heat of formation of naphthalene from benzene. This fails because of the resonance stabilisation energy in benzene, aromatic character.

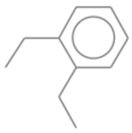
Example 2: 1,2,3,4-tetrahydronahpthalene

Let us take the example of 1,2,3,4-tetrahydronahpthalene



for which the experimental value according to Pedley et al. reads +26±2.0 kJ/mol.

You may start from 1,2-diethylbenzene for which the HGC model value reads -23 kJ/mol (experimental value -19 kJ/mol)



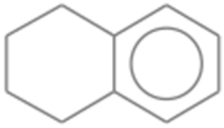
We now subtract 2 CH_3 Groups and add two CH_2 Groups resulting in -23 -2*(-42.36) +2* (-20.63) = 20.5 kJ/mol. When we now add 5.7 kJ/mol ring strain for cyclohexene we obtain 26.2 kJ/mol.

We may also choose to start from benzene (experimental value as it is a Group by itself in our GC model) and add 4 CH_2 Groups and add the substitution parameters value for di-alkyl-substituted benzenes (+18.5 kJ/mol) resulting in +83+4*(-20.63) + 18.5 (value AA3 in the Excel file) = 19 kJ/mol. When we add 5.7 kJ/mol ring strain for cyclohexene we arrive at 24.7 kJ/mol.

Both values, +26.2 and 24.7 kJ/mol, are within chemical accuracy when compared to the recommended experimental value of +26 ± 2.0 kJ/mol.

1,2,3,4,4a,5,6,7,8,8a-decahydronaphthalene

A less straightforward example is 1,2,3,4-tetrahydronaphthalene,



but it illustrates how one can 'push' the method to such a species which did not have a close equivalent in the parametrisation.
From the works by Rossini c.s. we have an experimental heat of formation (gas-phase) for 1,2-diethylbenzene of -19 kJ/mol. To arrive at the structure of 1,2,3,4-tetrahydronaphthalene we need to subtract 2 CH_3 Groups and subsequently add 2 CH_2 Groups. This leads to -19 -2* (-42.36) + 2*(-20.63) = +24.5 kJ/mol. This is to be compared with the experimental value of 26 ± 2.0 kJ/mol as reanalysed by Pedley, Naylor, et al., 1986, and these values agree within chemical accuracy. The G4 value reported by van der Spoel c.s. reads 21.3 kJ/mol (references in the full paper).