

Supporting Information

Atom Pair Contribution method: fast and general procedure to predict molecular formation enthalpies

Didier Mathieu
CEA, DAM, Le Ripault, 37260 Monts (France)

E-mail: didier.mathieu@cea.fr
Phone: +33-02-47-34-41-85

S1. Back-of-the-envelope evaluation of the formation enthalpy of nitrofuroxan

As a simple example illustrating evaluation by hand of formation enthalpies using the APC model, we consider the nitrofuroxan compound shown on Fig. S1.

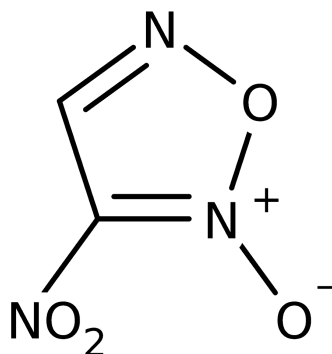


Figure S1: Molecular structure of nitrofuroxan.

The empirical formula of this compound is $C_2HN_3O_4$. The formation enthalpies of the isolated atoms, that is to say the standard enthalpies of atomization of the corresponding pure elements, are most conveniently picked up from Internet, for instance from the WebElements website (https://www.webelements.com/periodicity/enthalpy_atomisation/).

The values for C, H, N and O in kJ/mol are respectively 717, 218, 473 and 249. Therefore, the contribution to $\Delta_f H^0$ of the formation enthalpies of isolated atoms is:

$$\Delta_f H^0(ia) = 2 \times 717 + 1 \times 218 + 3 \times 473 + 4 \times 249 = 4067 \text{ kJ/mol} \quad (1)$$

The contribution of the ten covalent bonds can be split into four distinct contributions for single, aromatic, double and dative bonds, respectively, keeping in mind that the nitro group is viewed as a nitrogen atom with two oxygen neighbors attached through double bonds. Therefore, the total contribution of the two genuine single bonds in the molecule (excluding the dative bond), ie. respectively C–H and C–N, is obtained as:

$$\Delta_f H^0(1) = -415.5 - 376.4 = -791.9 \text{ kJ/mol} \quad (2)$$

The contribution of the five aromatic bonds is the sum of the contribution of the C:C bond, the two C:N bonds and the two N:O bonds, ie. respectively:

$$\Delta_f H^0(a) = -565.2 - 2 \times 493.1 - 2 \times 355.0 = -2261.4 \text{ kJ/mol} \quad (3)$$

The contribution of the two N=O bonds in the nitro group is:

$$\Delta_f H^0(2) = -2 \times 559.0 = -1118.0 \text{ kJ/mol} \quad (4)$$

The contribution of the N→O dative bond is simply:

$$\Delta_f H^0(d) = -328.5 \text{ kJ/mol} \quad (5)$$

Finally, the total contribution of all covalent bonds is:

$$\Delta_f H^0(\text{cov}) = -791.9 - 2261.4 - 1118.0 - 328.5 = -4499.8 \text{ kJ/mol} \quad (6)$$

The total repulsive contribution of geminal interactions arises from the addition of six terms corresponding respectively to C..H, C..N, C..O, N..H, N..N and O..O interactions:

$$\Delta_f H^0(\text{gem}) = 14.9 + 3 \times 23.4 + 5 \times 37.3 + 17.7 + 2 \times 16.1 + 2 \times 18.1 = 357.7 \text{ kJ/mol} \quad (7)$$

Finally, two corrections associated respectively with the 5-membered aromatic ring (R_{5a}) and the nitro group (NO_2) must be considered:

$$\Delta_f H^0(\text{cor}) = 53.4 + 212.6 = 266.0 \text{ kJ/mol} \quad (8)$$

The formation enthalpy is finally obtained as the sum of the four contributions associated respectively with gaseous atoms, covalent bonds, geminal interactions and structural corrections:

$$\Delta_f H^0 = 4067 - 4499.8 + 357.7 + 266.0 = 191 \text{ kJ/mol} \quad (9)$$

For comparison, the formation enthalpy of this compound was found to be about 221 kJ/mol using B3LYP/6-31G(d,p) calculations and isodesmic reactions, as detailed by Zhu et al. in Struct. Chem. 22, 149-159 (2011).

S2. Application of the accompanying script

APC calculations are even easier using the script `formation_enthalpy.py.txt` that accompanies the present article. Let us assume for instance that we are interested in the 15 molecules studied recently by Dorofeeva, Ryzhova and Sindtiskii as reported in Structural Chemistry, 26, pp. 1629-1640 (2015). These 15 compounds consist in guanidine and its amino and nitro derivatives. They may be encoded into SMILES strings as follows:

```

C=N
N[N+] (=O) [O-]
N=C(N)N
NN=C(N)N
NN=C(N)NN
NN=C(NN)NN
NC(N)=N[N+] (=O) [O-]
CNC(N)=N[N+] (=O) [O-]
CCNC(N)=N[N+] (=O) [O-]
NNC(N)=N[N+] (=O) [O-]
NNC(=N[N+] (=O) [O-])NN
NC(=N[N+] (=O) [O-])N[N+] (=O) [O-]
CN(C(N)=N[N+] (=O) [O-])[N+] (=O) [O-]
NC(N)=Nc1c([N+] (=O) [O-])cc([N+] (=O) [O-])cc1[N+] (=O) [O-]
NC(=N[N+] (=O) [O-])Nc1nnc(NC(N)=N[N+] (=O) [O-])nn1

```

The gas-phase formation enthalpies of these compounds have been very carefully evaluated by the authors of this previous study, using for every molecule a combination of various isodesmic reactions schemes and the state-of-the-art G4 thermochemical recipe (except for the two latter compounds for which G4(MP2) was used instead).

Using quantum chemical methods, such a study is made even more cumbersome as one needs to determine the lowest energy conformations for every structural formula. In addition to rotation about single bonds, one must in principle consider planar as well as slightly pyramidal configurations for every conjugated amine nitrogen.

In contrast, the formation enthalpies may be straightforwardly evaluated using the APC model, especially if the script `formation_enthalpy.py.txt` is used. Running this script from a terminal, the user just have to enter each compound as a SMILES string (or alternatively as an MDL .mol file for users unfamiliar with the SMILES notation). The program immediately displays the corresponding formation enthalpy in kJ/mol:

```

~/didier/APC> python formation_enthalpy.py.txt
C=N
57.8
N[N+] (=O) [O-]
17.7
N=C(N)N
27.3
NN=C(N)N
126.2
NN=C(N)NN
242.8
...

```

In practice, it is of course more convenient to prepare a file `smiles.inp` with the SMILES strings, and to save the results in an output file, as follows:

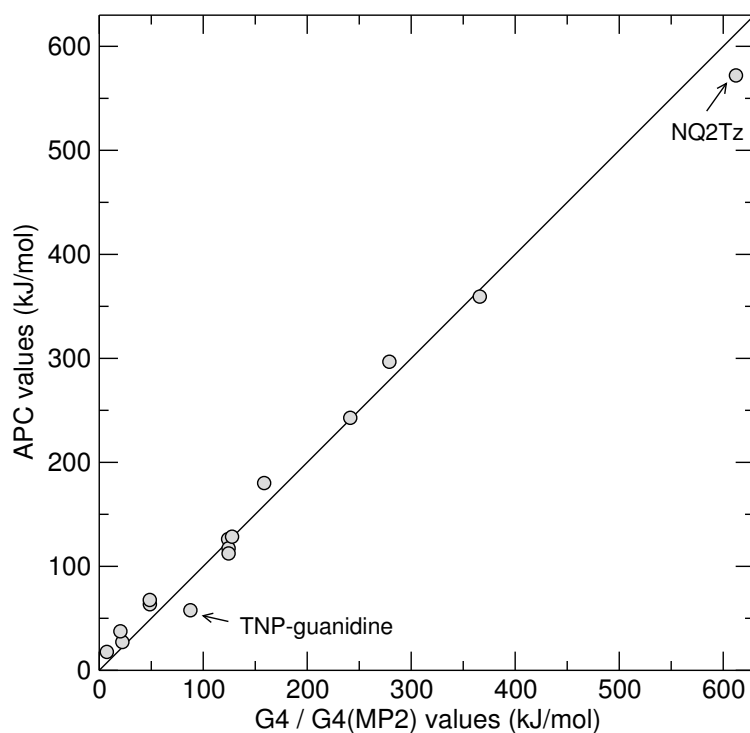


Figure S2: APC formation enthalpies of guanidine and its amino and nitro derivatives, plotted against G4 / G4(MP2) values

```
~/didier/APC> python formation_enthalpy.py.txt < smiles.inp > smiles.out
```

For this dataset, APC provides $\Delta_f H^0$ values with a root mean square deviation from G4/G4(MP2) values of 17 kJ/mol, and an average absolute deviation of 14 kJ/mol. The most significant deviations (−40 and −30 kJ/mol, respectively) are observed for NQ2Tz and TNP-guanidine, ie. the two largest compounds for which the ab initio calculations were carried out at the G4(MP2) level are associated with the largest uncertainties (namely ± 10.0 kJ/mol). The overall results are shown on Fig. S2.