# Supporting Information: Reliable prediction of association (free) energies of supramolecular complexes with heavy main group elements - the HS13L benchmark set

Johannes Gorges, Stefan Grimme, and Andreas Hansen*

*Mulliken Center for Theoretical Chemistry, Institute for Physical and Theoretical Chemistry, University of Bonn, Beringstr. 4, 53115 Bonn, Germany*

E-mail: hansen@thch.uni-bonn.de

## S1 Statistical error measures

Statistical measures for a set of data points $x_1, \cdots, x_n$ with references $r_1, \cdots, r_n$ are:

- Average:

$$\overline{x} = \frac{1}{n}\sum_{i}^{n} x_i$$

- Mean deviation(MD):

$$MD = \frac{1}{n}\sum_{i}^{n}(x_i - r_i)$$

- Mean absolute deviation(MAD):

$$MAD = \frac{1}{n}\sum_{i}^{n}|x_i - r_i|$$

- relative Mean absolute deviation(relMAD):

$$relMAD = \frac{1}{n}\sum_{i}^{n}\frac{|x_i - r_i|}{|r_i|}$$

- standard deviation (SD):

$$SD = \sqrt{\frac{1}{n-1}\sum_{i}^{n}(x_i - r_i - MD)^2}$$

## S2 Revised D4 parameter

In the development of another project in our group, revised D4 parameter were developed for $\omega$B97X.[S1] The respective parameter are given in the following Table.

Table S1: Revised D4 parameter for the $\omega$B97X DFA. Details can be found in the respective publication.[S1]

| s6 | s8 | s9 | a1 | a2 |
|---|---|---|---|---|
| 1.0 | 0.4485 | 1.0 | 0.3306 | 4.2790 |

# S3 Convergence Problems

Due to convergence problems in the geometry optimization and frequency calculation of the guest of complex **11**, the respective calculations with GFN-FF and GFN2 were performed with the GBSA solvation model instead of the ALPB solvation model. In the SPH approach for GFN2-xTB, this problem did not occur. For the same reason, PM6-D3H4X frequencies for **11**_CI could not be obtained and the respective values calculated for **11** were taken instead.

# S4 Effect of the solvent model in the geometry optimization

Table S2: root-mean- square deviation (RMSD) of the heavy atoms of host, guest, and complex of **8** of r$^2$SCAN-3c optimized geometries with SMD, COSMO and CPCM solvation model with respect to r$^2$SCAN-3c[DCOSMO-RS] optimized geometries in Å. Association free energies $\Delta G_\mathrm{a}$ computed with the default workflow (r$^2$SCAN-3c + GFN2-xTB-SPH + COSMO-RS(16) normal) on the respective geometries are given in kcal mol$^{-1}$.

| | SMD | COSMO | CPCM | DCOSMO-RS |
|---|---|---|---|---|
| RMSD(host) | 4.00E-02 | 4.76E-02 | 4.90E-02 | - |
| RMSD(guest) | 7.06E-04 | 2.77E-04 | 2.77E-04 | - |
| RMSD(complex) | 1.12E-01 | 5.21E-02 | 9.12E-01 | - |
| $\Delta G$ | -4.91 | -5.31 | -5.32 | -5.44 |

# S5 Influence of conformers

For the default workflow (r$^2$SCAN-3c + GFN2-xTB-SPH + COSMO-RS(16) normal) the ensemble free energy $\overline{G}$ was computed as Boltzmann average from all remaining conformers in the 3.5 kcal mol$^{-1}$window after *Part2* of CENSO.

$$G_\mathrm{Bw} = \sum_i^N p_i G_i$$

$$p_i = \frac{e^{-G_i/\mathrm{k}T}}{\sum_i^N e^{-G_i/\mathrm{k}T}}$$

Table S3: Experimental and calculated association free energies at the default level of theory ($\mathrm{r^2}$SCAN-3c energies, GFN2-xTB[ALPB]-SPH thermostatistical contributions, and COSMO-RS(16)-normal($\mathrm{r^2}$SCAN-3c) solvation free energies) for the lowest conformer of each complex of the HS13L set, as well as Boltzmann-weighted average $\Delta G_{\mathrm{Bw}}$ of all conformers found below the threshold of 3.5 kcal mol$^{-1}$.

| complex | $\Delta G_{\mathrm{EXP}}$ | $\Delta G_{\mathrm{calc}}$ | $\Delta G_{\mathrm{Bw}}$ |
|---|---|---|---|
| 1 | -8.4 | -7.2 | -7.2 |
| 2 | -5.8 | -4.4 | -4.5 |
| 3 | -4.7 | -12.5 | -12.4 |
| 4 | -5.7 | -5.1 | -5.2 |
| 5 | -1.9 | -4.5 | -4.5 |
| 6 | -6.8 | -1.9 | -1.6 |
| 7 | -6.7 | -4.0 | -4.1 |
| 8 | -5.0 | -4.9 | -4.8 |
| 9 | -5.5 | -4.7 | -4.7 |
| 10 | -9.2 | -5.4 | -5.1 |
| 11 | -8.1 | -12.2 | -12.3 |
| 12 | -5.2 | 1.8 | 1.1 |
| 13 | -7.5 | -2.5 | -3.0 |

# S6 Counterpoise correction

Table S4: Boys Bernardi Counterpoise correction $E_{\mathrm{cp}}$ in kcal mol$^{-1}$for complex **1** for best-performing DFAs of each class with respect to the backcorrected experimental values.

| DFA | $E_{\mathrm{CP}}$ / kcal mol$^{-1}$ |
|---|---|
| PBE-NL | 0.08 |
| B97M-V | 0.03 |
| PW6B95-NL | 0.15 |
| $\omega$B97X-V | 0.02 |
| PWPB95-NL | 1.25 |

# S7 Magnitude of the ATM term

Table S5: Magnitude of the non-additive ATM term contribution (which is the same for each DFA) in kcal mol$^{-1}$to the association energy for each complex of HS13L and HS13L-CI. The CI denotes that the complex was docked with counterions to neutralize its global charge.

| **complex** | **ATM** / kcal mol$^{-1}$ |
|---|---|
| **1** | 1.11 |
| **2** | 1.95 |
| **3** | 3.93 |
| **4** | 2.86 |
| **5** | -0.02 |
| **6** | 3.42 |
| **7** | 2.76 |
| **8** | 1.27 |
| **9** | 1.33 |
| **10** | 2.91 |
| **11** | 1.45 |
| **12** | 1.55 |
| **13** | 1.30 |
| **10_CI** | 3.00 |
| **11_CI** | 1.34 |
| **12_CI** | 1.62 |
| **13_CI** | 1.55 |

# S8 DFT frequencies

Table S6: Comparison of $\Delta G_{\mathrm{mRRHO}}$ obtained with SQM methods to those obtained from r$^2$SCAN-3c-SPH calculations for the 5 smallest complexes of the HS13L. The MD and MAD is given with respect to r$^2$SCAN-3c-SPH. Note that GFN1-xTB, GFN2-xTB, and GFN-FF were used with the ALPB implicit solvation model and PM6 and PM7 with the COSMO implicit solvation model.

| **complex** | **GFN-FF** | **GFN1** | **GFN2** | **PM6]** | **PM7** | **GFN2-xTB-SPH** | **r2scan-3c** |
|---|---|---|---|---|---|---|---|
| **1** | 13.90 | 10.57 | 13.54 | 12.70 | 12.52 | 11.67 | 9.98 |
| **2** | 16.46 | 16.31 | 15.76 | 16.35 | 16.67 | 15.69 | 16.10 |
| **5** | 11.58 | 11.73 | 11.15 | 11.87 | 16.09 | 11.89 | 13.06 |
| **8** | 14.14 | 14.05 | 13.71 | 13.96 | 14.75 | 14.29 | 14.37 |
| **9** | 14.83 | 14.48 | 14.62 | 14.01 | 15.53 | 14.73 | 13.94 |
| **MD** | **0.6** | **-0.2** | **0.2** | **0.3** | **1.6** | **0.0** | |
| **MAD** | **1.5** | **0.6** | **1.6** | **1.1** | **1.6** | **0.8** | |

Table S7: Symmetry numbers for optimized geometries of the lowest conformer of host, guest, and complex, respectively of the HS13L. Charged systems which were attributed with counterions (**10_B**, **10_AB**, **11_B**, **11_AB**, **12_B**, **12_AB**, **13_B**, **13_AB**) in the HS13L-CI have symmetry number 1.

| complex | A | B | AB |
|---|---|---|---|
| **1** | 12 | 2 | 2 |
| **2** | 1 | 4 | 1 |
| **3** | 1 | 10 | 1 |
| **4** | 1 | 1 | 1 |
| **5** | 3 | 2 | 1 |
| **6** | 1 | 3 | 1 |
| **7** | 14 | 1 | 1 |
| **8** | 2 | 2 | 1 |
| **9** | 2 | 2 | 1 |
| **10** | 12 | 1 | 1 |
| **11** | 2 | 60 | 1 |
| **12** | 1 | 1 | 1 |
| **13** | 1 | 1 | 1 |
| **10_CI** | 12 | 1 | 1 |
| **11_CI** | 2 | 1 | 1 |
| **12_CI** | 1 | 1 | 1 |
| **13_CI** | 1 | 1 | 1 |

# S9 Statistical evaluation of HS13L without counterions

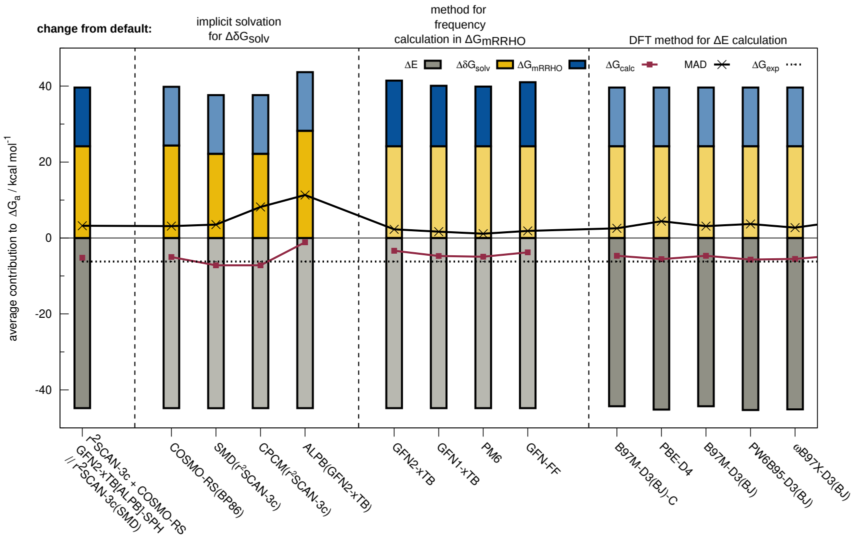

Figure S1: Contributions to the calculated $\Delta G_{\text{calc}}$ averaged over all complexes of the HS13L set. The leftmost bar illustrates the default approach used in this work, while the others illustrate the effect of using a different model or level of theory for the calculation of the solvation contribution, the thermostatistical correction and the electronic energy. Contributions that are not affected by these variations are depicted in brighter colors.

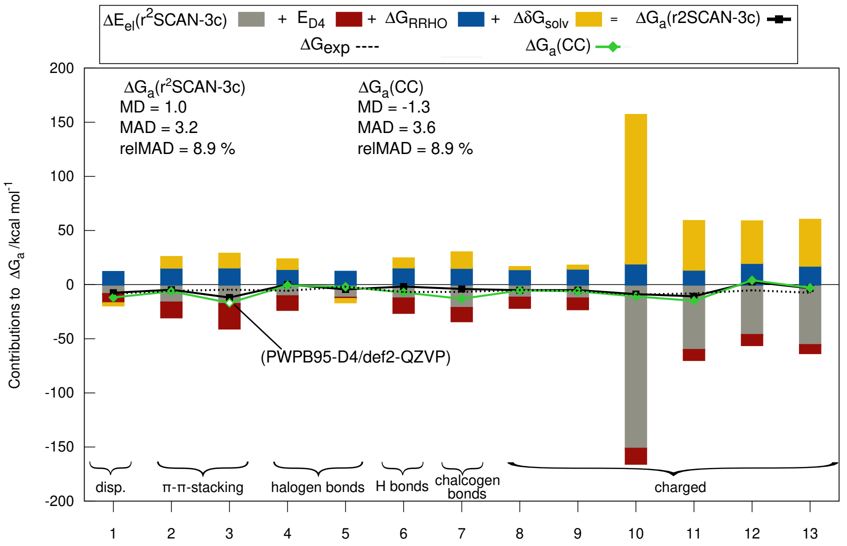

Figure S2: Contributions to the association free energy $\Delta G_a$ for each complex of the HS13L in kcal mol$^{-1}$. MD, MAD, and relMAD are also given. For comparison, $\Delta G_a$ when using DLPNO-CCSD(T1)/CBS instead of $r^2$SCAN-3c are also given. For complex **3** only PWPB95-D4/def2-QZVP values could be obtained and are shown instead.

Table S8: MDs, MADs and SDs in kcal mol$^{-1}$to the DLPNO-CCSD(T1)/CBS reference values of all assessed DFAs in def2-QZVP basis in combination with their best dispersion correction and composite methods for the HS13L. For complex **3** PWPB95-D4/def2-QZVP values were used as reference, as no DLPNO-CCSD(T1)/CBS reference value could be obtained for this complex.*:downscaled s9.

| | **MD** | **MAD** | **SD** |
|---:|---:|---:|---:|
| **r$^2$SCAN-3c** | 2.27 | 2.91 | 3.19 |
| **PBEh-3c** | 5.92 | 6.81 | 7.91 |
| **B97-3c** | 2.86 | 4.67 | 9.95 |
| **B97M-V-C** | -1.31 | 2.87 | 3.45 |
| **PBE-NL** | -0.42 | 2.81 | 3.63 |
| **RPBE-NL** | -2.22 | 3.25 | 3.55 |
| **R$^2$SCAN-D4** | 1.63 | 2.39 | 3.16 |
| **M06L-D3(0)** | -1.52 | 3.78 | 4.85 |
| **B97M-V** | -0.77 | 1.90 | 2.57 |
| **PW6B95-NL** | 0.05 | 2.46 | 3.55 |
| **B3LYP-D4** | -1.43 | 2.65 | 2.78 |
| **M06-2X-D3(0)** | 1.23 | 3.13 | 4.27 |
| **PBE0-D4** | 1.22 | 2.30 | 3.35 |
| **$\omega$B97X-V** | -1.43 | 2.87 | 3.32 |
| **$\omega$B97M-V** | -2.54 | 3.47 | 3.18 |
| **PWPB95-D3** | 0.15 | 2.30 | 3.03 |
| **rev-DSDPBEP86-D4** | 2.86 | 3.77 | 3.74 |
| **rev-DSDPBEP86-D4$^*$** | 0.51 | 2.62 | 4.08 |
| **MP2-CBS** | -13.1 | 13.1 | 12.4 |

Table S9: Mean deviation (MD), mean absolute deviation (MAD), and standard deviation (SD) to the DLPNO-CCSD(T1)/CBS reference values of tested semiempirical methods for the HS13L in kcal mol$^{-1}$. The average root-mean square deviation of the heavy atom positions ($\overline{RMSD}$) between the with implicit solvation (ALPB for the GFN methods and COSMO for the PM methods) optimized geometries with respect to the r$^2$SCAN-3c optimized geometries for the HS13L-CI are given in Å.

| | **MD** | **MAD** | **SD** | $\overline{RMSD}$ |
|---:|---:|---:|---:|---:|
| **GFN1-xTB** | 2.7 | 6.1 | 7.0 | 0.28 |
| **GFN2-xTB** | 6.7 | 7.4 | 8.2 | 0.27 |
| **GFN-FF** | 4.0 | 5.7 | 6.3 | 0.30 |
| **PM6** | -0.3 | 4.5 | 6.1 | 0.29 |
| **PM7** | -5.0 | 11.3 | 12.4 | 0.32 |

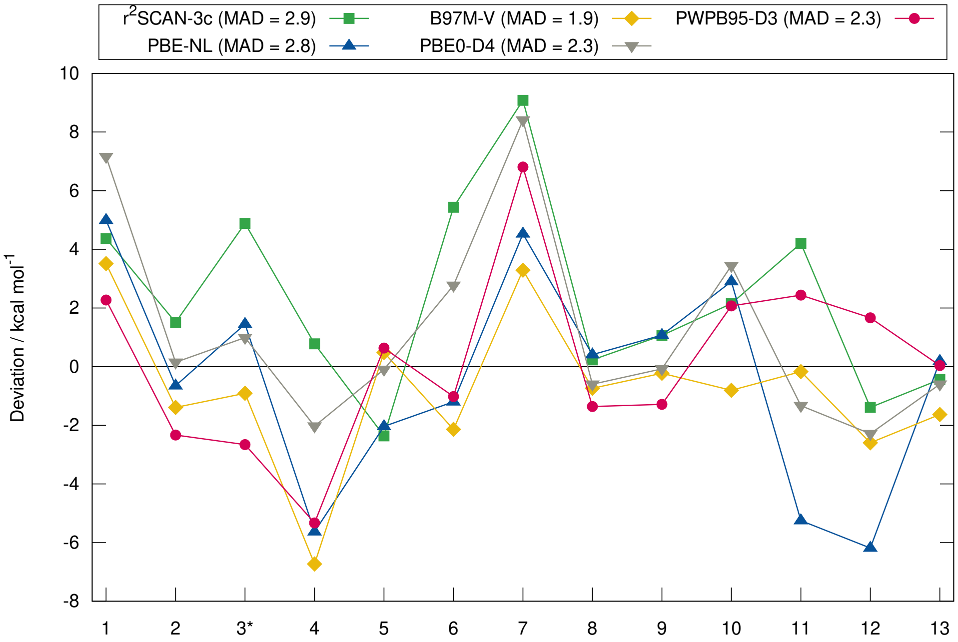

Figure S3: Deviations in kcal mol$^{-1}$ to the DLPNO-CCSD(T1)/CBS reference values of best-performing methods of each class for HS13L. A negative deviation indicates overbinding, a positive underbinding. For complex **3** PWPB95-D4/def2-QZVP is used as reference, as no DLPNO-CCSD(T1)/CBS reference value could be obtained for this complex indicated by the *.

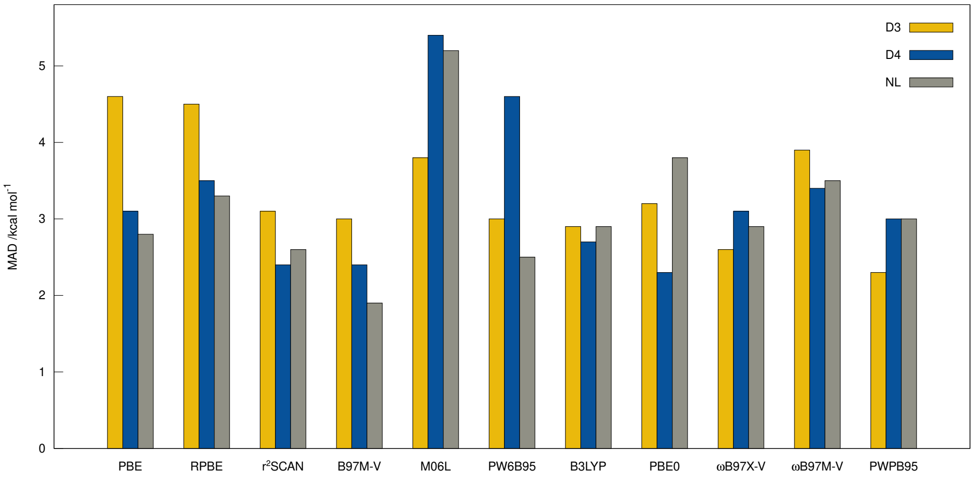

Figure S4: Mean absolute Deviations (MADs) to the DLPNO-CCSD(T1)/CBS reference values of assessed DFAs for the HS13L in combination with D3, D4, and the nonlocal VV10 dispersion correction.

# References

(S1) M. Müller, A. Hansen, S. Grimme, "$\omega$B97X-3c: A composite range-separated hybrid DFT method with a molecule-optimized polarized valence double-$\zeta$ basis set", submitted to *J. Chem. Phys.* (2022).