

Quantifying the accuracy of benchmark calculations of non-covalent complexes

Jan Řezáč

Institute of Organic Chemistry and Biochemistry, AS CR

November 30, 2015

Non-covalent interactions

- Non-covalent interactions are important in many areas of chemistry
- Difficult to describe by computational methods
- Accurate treatment of electron correlation needed
- **Accuracy of computational methods not a priori known**

Quantifying the accuracy of computational methods

- Experiment - dissociation energy D_0 is complex quantity (ΔH_0^0)
- Hard to reproduce by calculations
- Limited experimental data

- Comparing calculation to calculation is straightforward
- Interaction energy (ΔE) easy to calculate
- Reliable benchmark needed – an order of magnitude more accurate than tested methods

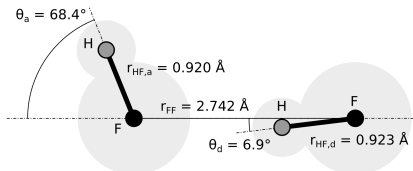
Experimental benchmark

- D_0 can not be calculated directly
- Calculated components are not observables
- Structure \rightarrow deformation energy E^{def}
- Nuclear dynamics / Vibrational analysis $\rightarrow \Delta ZPVE$
- Interaction energy ΔE
- Choice of method (coverage of correlation)
- Convergence with basis set
- Other effects/approximations

Experimental benchmark – HF dimer

- Experiment¹ with accuracy $\pm 1\text{cm}^{-1}$
- Small enough to apply the best quantum chemistry offers²
- Uncertainty of components due to basis set size

Dissociation energy components (kcal/mol)		
Uncertainty of the geometry		± 0.003
HF energy	-3.822	± 0.000
CCSD(T) correlation energy	-0.786	± 0.003
$\Delta\text{CCSDT(Q)}$	-0.009	± 0.003
ΔCCSDTQ	0.001	± 0.001
Relativistic effects	0.016	± 0.001
Diagonal B-O correction	-0.012	± 0.000
Deformation energy	0.028	± 0.001
Harmonic ZPVE	1.805	± 0.017
ZPVE anharmonicity (VPT2)	-0.185	± 0.019
Sum – calculated D0	-2.964	± 0.047
Experimental D0	-3.036	± 0.003



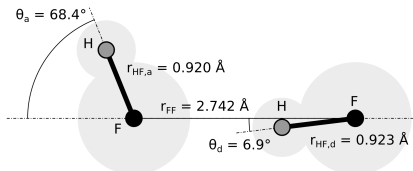
¹Bohac, E. J.; Marshall, M. D.; Miller, R. E. *J. Chem. Phys.* 1992, 96 (9), 6681–6695.

²Řezáč, J.; Hobza, P. *J. Chem. Theory Comput.* 2014, 10 (8), 3066–3073.

Experimental benchmark – HF dimer

- Experiment¹ with accuracy $\pm 1\text{cm}^{-1}$
- Small enough to apply the best quantum chemistry offers²
- Uncertainty of components due to basis set size

Dissociation energy components (kcal/mol)		
Uncertainty of the geometry		± 0.003
HF energy	-3.822	± 0.000
CCSD(T) correlation energy	-0.786	± 0.003
$\Delta\text{CCSDT(Q)}$	-0.009	± 0.003
ΔCCSDTQ	0.001	± 0.001
Relativistic effects	0.016	± 0.001
Diagonal B-O correction	-0.012	± 0.000
Deformation energy	0.028	± 0.001
Harmonic ZPVE	1.805	± 0.017
ZPVE anharmonicity (VPT2)	-0.185	± 0.019
Sum – calculated D0	-2.964	± 0.047
Experimental D0	-3.036	± 0.003



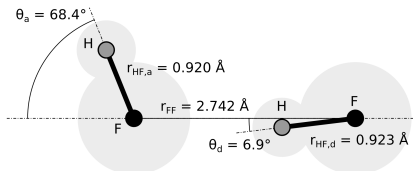
¹Bohac, E. J.; Marshall, M. D.; Miller, R. E. J. Chem. Phys. 1992, 96 (9), 6681–6695.

²Řezáč, J.; Hobza, P. J. Chem. Theory Comput. 2014, 10 (8), 3066–3073.

Experimental benchmark – HF dimer

- Experiment¹ with accuracy $\pm 1\text{cm}^{-1}$
- Small enough to apply the best quantum chemistry offers²
- Uncertainty of components due to basis set size

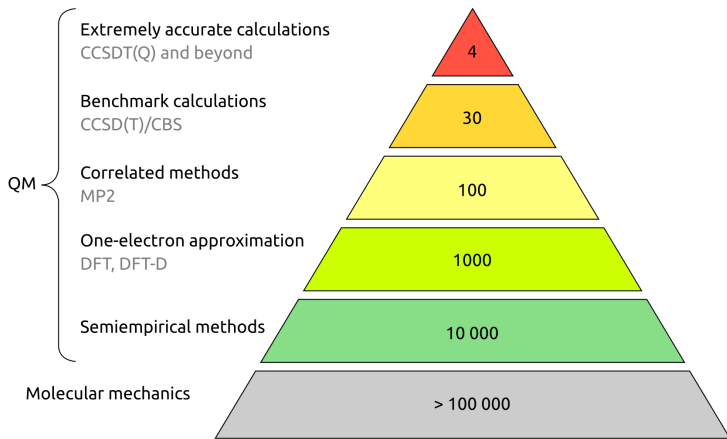
Dissociation energy components (kcal/mol)		
Uncertainty of the geometry		± 0.003
HF energy	-3.822	± 0.000
CCSD(T) correlation energy	-0.786	± 0.003
$\Delta\text{CCSDT(Q)}$	-0.009	± 0.003
ΔCCSDTQ	0.001	± 0.001
Relativistic effects	0.016	± 0.001
Diagonal B-O correction	-0.012	± 0.000
Deformation energy	0.028	± 0.001
Harmonic ZPVE	1.805	± 0.017
ZPVE anharmonicity (VPT2)	-0.185	± 0.019
Sum – calculated D0	-2.964	± 0.047
Experimental D0	-3.036	± 0.003



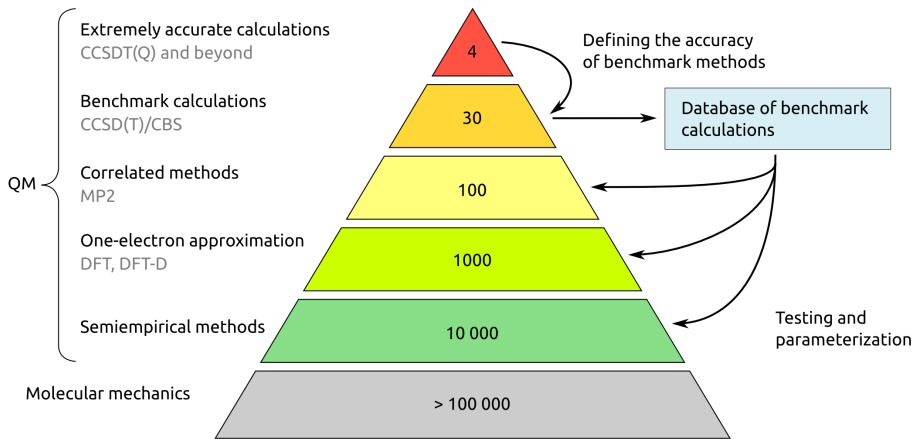
¹Bohac, E. J.; Marshall, M. D.; Miller, R. E. J. Chem. Phys. 1992, 96 (9), 6681–6695.

²Řezáč, J.; Hobza, P. J. Chem. Theory Comput. 2014, 10 (8), 3066–3073.

Computational benchmarking



What do we do?



The "gold standard" in benchmarking: CCSD(T)/CBS

- "First method accurate enough, last computationally tractable"
- Small systems – direct extrapolation to CBS
- Larger systems – composite scheme

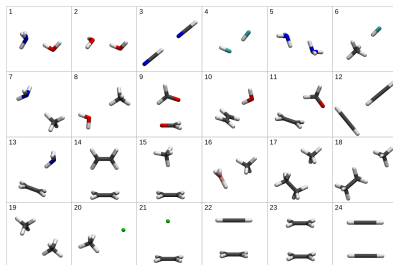
$$E^{CCSD(T)} = E^{HF} + E^{MP2} + \delta E^{CCSD(T)}$$

$$\delta E^{CCSD(T)} = E^{CCSD(T)} - E^{MP2}$$

- Counterpoise correction, frozen core approximation
- **What is the accuracy of this benchmark?**

Benchmarking the CCSD(T)/CBS benchmark

- A24 data set - 24 complexes featuring diverse interactions³
- Best estimate of interaction energies:
CCSD(T)/CBS + Δ CCSDT(Q) + all electron and relativistic corrections
- Recently updated with more accurate CCSDT(Q) calculations⁴



³Řezáč, J.; Hobza, P. J. Chem. Theory Comput. 2013, 9 (5), 2151–2155.

⁴Řezáč, J.; Dubecký, M.; Jurečka, P.; Hobza, P. Phys. Chem. Chem. Phys. 2015, 17 (29), 19268–19277.

Errors inherent to CCSD(T)/CBS

- Errors inherent to frozen-core CCSD(T), based on updated reference from ⁵
 - Truncation of the CC series, **1.14%** vs. CCSDT(Q)
CCSDT(Q) covers about 95% of correlation missing in CCSD(T),⁶
Convergence of ΔE^{int} tested up to FCI⁷
 - Frozen core approximation, **0.57%**
 - Relativistic effects, **0.14%** (in light elements)
 - Born-Oppenheimer approximation, ca. **0.1%** (HF dimer)
- Overall error 1.7%

⁵Řezáč, J.; Dubecký, M.; Jurečka, P.; Hobza, P. Phys. Chem. Chem. Phys. 2015, 17 (29), 19268–19277.

⁶Bomble, Y. J.; Stanton, J. F.; Kállay, M.; Gauss, J. J. Chem. Phys. 2005, 123 (5), 054101–054101.

⁷Šimová, L.; Řezáč, J.; Hobza, P. J. Chem. Theory Comput. 2013, 9 (8), 3420–3428.


Summary I

- CCSD(T)/CBS has error of about 1.7% compared to exact solution
- Largest source of error is approximate treatment of correlation
- CCSDT(Q) is a solution applicable to small model systems

Basis set error in composite CCSD(T)/CBS

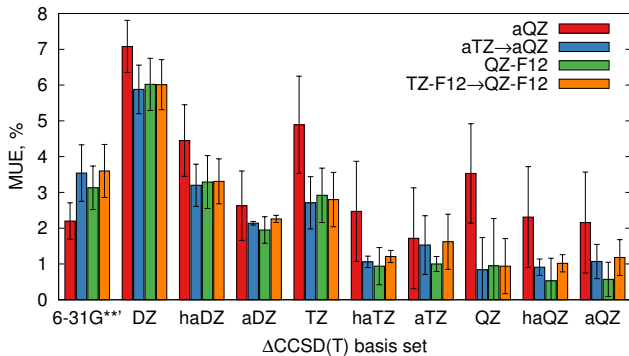
- Accuracy of composite schemes applicable to larger molecules
- Evaluated in the A24 set, reference is CCSD(T) extrapolated from aug-cc-pV(T,Q,5)Z basis sets
- Composite scheme $E^{CCSD(T)} = E^{HF} + E^{MP2} + \delta E^{CCSD(T)}$
 - E^{HF} in the same basis as MP2, not extrapolated
 - E^{MP2} – multiple ways to achieve CBS:
 - Single calculation
 - Extrapolation (Helgaker⁸, $\Delta E_X^{corr} = E_{CBS}^{corr} + aX^{-3}$)
 - Explicitly correlated (MP2-F12)
 - Explicitly correlated, extrapolated (Hill⁹)
 - $\delta E^{CCSD(T)}$ in basis no larger than MP2

⁸Helgaker, T.; Klopper, W.; Koch, H.; Noga, J. J. Chem. Phys. 1997, 106 (23), 9639–9646.

⁹Hill, J. G.; Peterson, K. A.; Knizia, G.; Werner, H.-J. J. Chem. Phys. 2009, 131, 194105. 

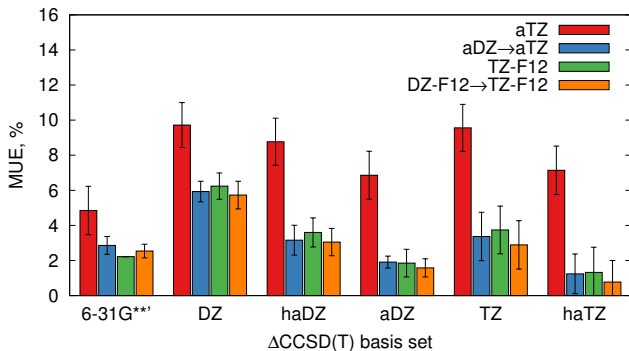
Basis set error in composite CCSD(T)/CBS

- MP2 in QZ



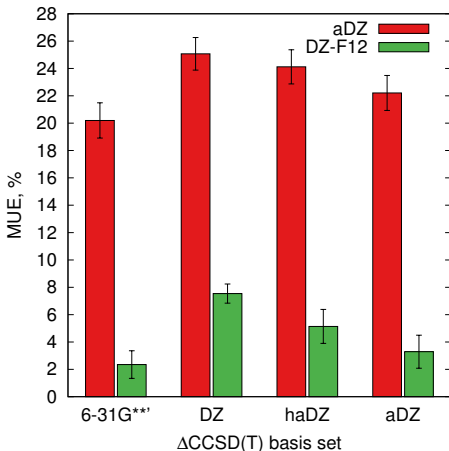
Basis set error in composite CCSD(T)/CBS

- MP2 in TZ



Basis set error in composite CCSD(T)/CBS

- MP2 in DZ



Summary II

- Extrapolation and MP2-F12 yield comparable results at comparable expense
- MP2-F12 is the only solution in small (DZ) basis
- The bottleneck is always the CCSD(T) correction
- 1% accuracy can be achieved with CCSD(T) correction in triple-zeta basis

Benchmark data sets and their accuracy

- S66^{10,11} and X40¹² use following scheme:
HF/aug-cc-pVQZ + MP2/aug-cc-pV(T→Q)Z +
ΔCCSD(T)/heavy-aug-cc-pVTZ
Accuracy of 1.1% in A24
- Non-equilibrium geometries in S66x8, S66a8 and X40x10:
HF/aug-cc-pVQZ + MP2/aug-cc-pV(T→Q)Z +
ΔCCSD(T)/aug-cc-pVDZ
Accuracy of 2.1% in A24

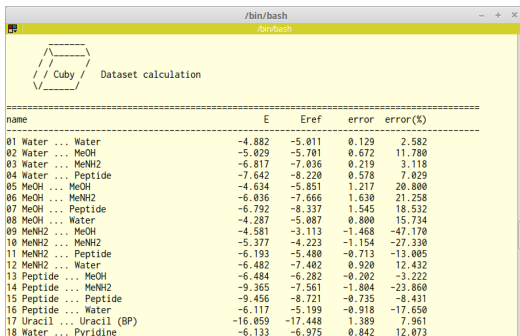
¹⁰Řezáč, J.; Riley, K. E.; Hobza, P. J. Chem. Theory Comput. 2011, 7 (8), 2427–2438.

¹¹Řezáč, J.; Riley, K. E.; Hobza, P. J. Chem. Theory Comput. 2011, 7 (11), 3466–3470.

¹²Řezáč, J.; Riley, K. E.; Hobza, P. J. Chem. Theory Comput. 2012, 8 (11), 4285–4292.

Automation of dataset calculations

- Cuby framework developed by the author¹³
- Automatic calculation of data set, processing of the results
- Implements also CC correction and extrapolation to CBS
- Data sets discussed here are distributed with the framework



name	E	Eref	error	error(%)
01 Water ... Water	-4.882	-5.011	0.129	2.582
02 Water ... MeOH	-5.029	-5.701	0.672	11.780
03 Water ... MeNH2	-6.817	-7.036	0.219	3.118
04 Water ... Peptide	-7.642	-8.220	0.578	7.029
05 MeOH ... MeOH	-4.634	-5.851	1.217	20.800
06 MeOH ... MeNH2	-6.036	-7.666	1.630	21.258
07 MeOH ... Peptide	-6.792	-8.337	1.545	18.532
08 MeOH ... Water	-4.287	-5.007	0.800	15.734
09 MeNH2 ... MeOH	-4.581	-3.113	-1.468	-47.170
10 MeNH2 ... MeNH2	-5.377	-4.223	-1.154	-27.330
11 MeNH2 ... Peptide	-6.193	-5.480	-0.713	-13.005
12 MeNH2 ... Water	-6.482	-7.402	0.920	12.432
13 Peptide ... MeOH	-6.484	-6.282	-0.202	-3.222
14 Peptide ... MeNH2	-9.365	-7.561	-1.804	-23.860
15 Peptide ... Peptide	-9.456	-8.721	-0.735	-8.431
16 Peptide ... Water	-6.117	-5.199	-0.918	-17.650
17 Uracil ... Uracil (BP)	-16.059	-17.448	1.389	7.961
18 Water ... Pyridine	-6.133	-6.975	0.842	12.073

¹³<http://cuby4.molecular.cz>

Automation of dataset calculations

- Cuby framework developed by the author¹³
- Automatic calculation of data set, processing of the results
- Implements also CC correction and extrapolation to CBS
- Data sets discussed here are distributed with the framework

```

/bin/bash
66 MeNH2 ... Pyridine          -4.297  -3.966  -0.331  -8.340
-----
RMSE          0.979  kcal/mol
MUE           0.788  kcal/mol
-----
MSE           0.162  kcal/mol
min          -1.804  kcal/mol
max           2.660  kcal/mol
range         4.464  kcal/mol
min abs       0.032  kcal/mol
max abs       2.660  kcal/mol
-----
RMSE          25.204  %
MUE           19.104  %
MSE           -0.390  %
min          -60.382  %
max           56.817  %
range        117.199  %
min abs       0.970  %
max abs       60.382  %
-----
MUE/|avg|     14.328  %
-----
H-bonds      (23) RMSE  1.089  MSE   0.327  kcal/mol
dispersion   (23) RMSE  0.854  MSE  -0.355  kcal/mol
others       (20) RMSE  0.981  MSE   0.566  kcal/mol
-----

```

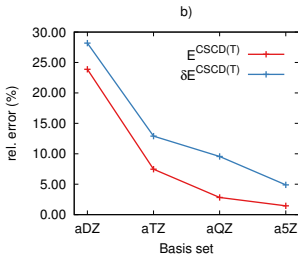
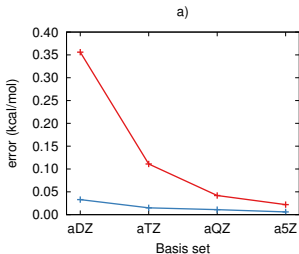
¹³<http://cuby4.molecular.cz>

Conclusions

- Accuracy of composite CSCD(T)/CBS calculations is now well characterized
- Excellent accuracy of 1–2% can be achieved in routine calculations of medium-sized molecules
- Going beyond that is extremely difficult

Why composite CCSD(T)/CBS works so well?

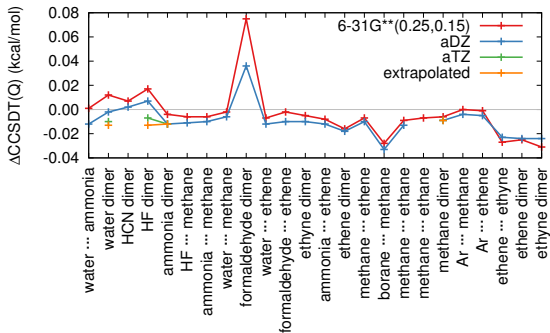
- Using composite scheme
 $MP2/CBS + (E^{CCSD(T)} - E^{MP2})/smaller\ basis$
 - Historically, developed as most economic option
 - No special convergence properties of the correction
 - However, its accuracy is hard to beat
- More expensive $CCSD/CBS + (T)/smaller\ basis$ is not better
- The same applies to $CCSD(T)-F12$ or $CCSD(T^*)-F12^{14}$



¹⁴Patkowski, K. The Journal of Chemical Physics 2012, 137 (3), 034103.

Basis set dependence of CCSDT(Q) correction

- $O(N^9)$ scaling limits basis set size
- Convergence differs with nature of the interaction^{15,16}



¹⁵Smith, D. G. A.; Jankowski, P.; Slawik, M.; Witek, H. A.; Patkowski, K. J. Chem. Theory Comput. 2014, 10 (8), 3140–3150.

¹⁶Demovičová, L.; Hobza, P.; Řezáč, J. Phys. Chem. Chem. Phys. 2014, 16 (36), 19115–19121.