Accurate description of non-covalent interactions in QM methods applicable to large systems

Jan Řezáč

Institute of Organic Chemistry and Biochemistry,
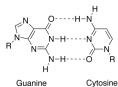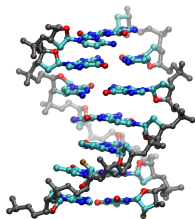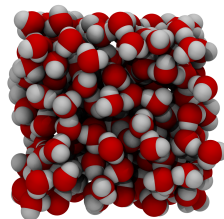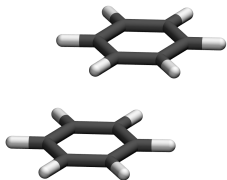Academy of Sciences of the Czech Republic,
Prague, Czech Republic

May 21, 2014

Importance of noncovalent interactions

Computational methods

Extremely accurate calculations post-CCSD(T)	
Benchmark calculations CCSD(T)/CBS	
Correlated methods MP2	
One-electron approximation DFT	
Semiempirical methods	
Molecular mechanics	

QM

4

30

100

1000

10 000

> 100 000

Application to biomolecules:

Building blocks and their interactions

Active site of a protein

Whole protein

Protein in environment
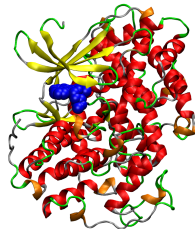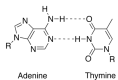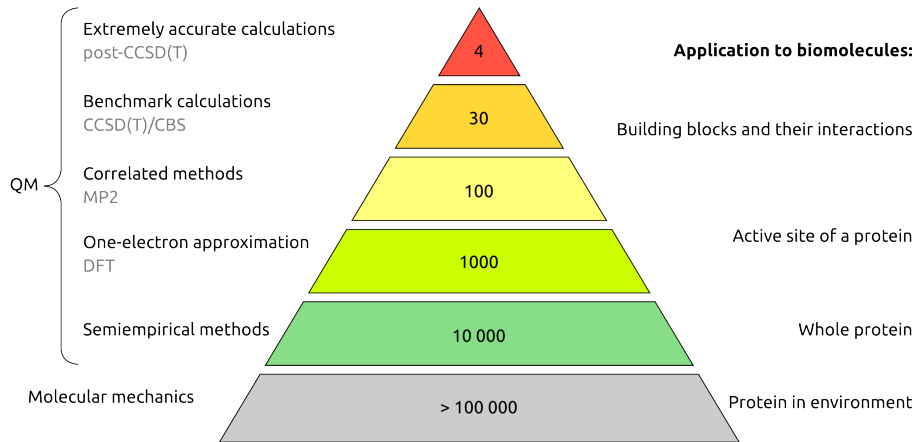
What do we do?
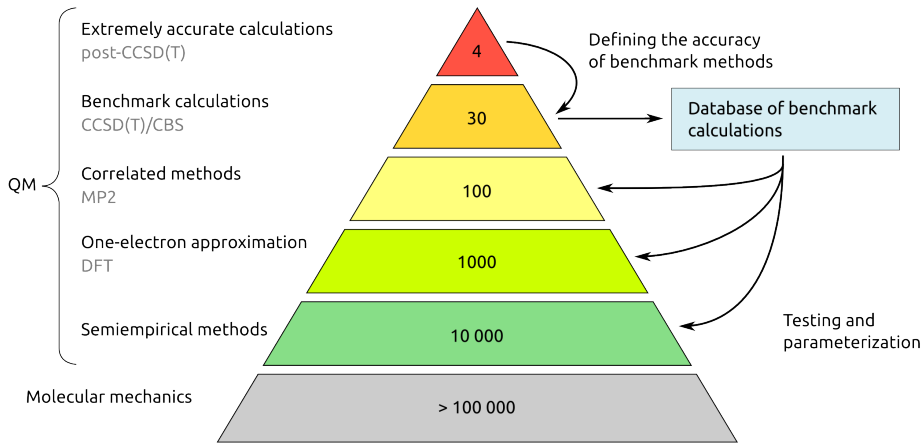
Benchmarks

The benchmark method: CCSD(T)/CBS

- Accurate and consistent description of different interactions
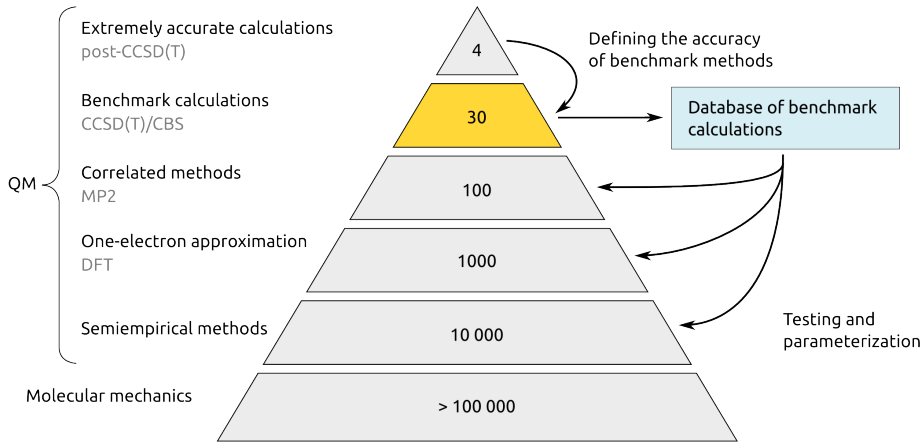- Applicable to reasonably large systems (up to cca. 40 atoms)
- Composite calculation:
  $E = E^{MP2/CBS} + \Delta CCSD(T)$
  $\Delta CCSD(T) = E^{CCSD(T)/BS} - E^{MP2/BS}$
- BS = aug-cc-pVDZ $\rightarrow$ error 3%
  BS = aug-cc-pVTZ $\rightarrow$ error 1%

S66 data set

- **S66** - 66 comlexes of (bio)organic molecules[1,2]
- All important interaction motifs covered
- Intended as a replacement of S22

---
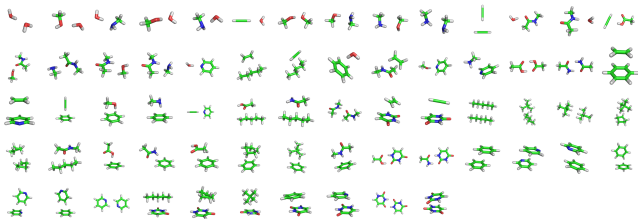
[1]Řezáč, J.; Riley, K. E.; Hobza, P. J. Chem. Theory Comput. 2011, 7, 2427–2438.

[2]Řezáč, J.; Riley, K. E.; Hobza, P. J. Chem. Theory Comput. 2011, 7, 3466–3470.

S66 data set

- **S66** - 66 comlexes of (bio)organic molecules[1,2]
- All important interaction motifs covered
- Intended as a replacement of S22

- Auxiliary data sets covering nonequilibrium geometries:
- **S66x8** - Dissociation curves
  S66a8 - Angular displacements
- More than 1000 CSCD(T)/CBS points

---

[1]Řezáč, J.; Riley, K. E.; Hobza, P. J. Chem. Theory Comput. 2011, 7, 2427–2438.

[2]Řezáč, J.; Riley, K. E.; Hobza, P. J. Chem. Theory Comput. 2011, 7, 3466–3470.
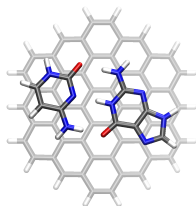
Other benchmark data sets

Halogented molecules

- **X40** - 40 systems, including halogen bonding[3]
- **X40x10** - dissociation curves

Large complexes

- **L7** - large systems, up to 110 atoms[4]

---
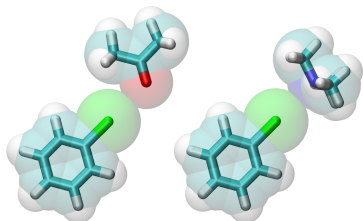
[3]Řezáč, J.; Riley, K. E.; Hobza, P. J. Chem. Theory Comput. 2012, 8, 4285–4292.

[4]Sedlák, R.; Janowski, T.; Pitoňák, M.; Řezáč, J.; Pulay, P.; Hobza, P. J. Chem. Theory Comput. 2013

Jan Řezáč (IOCB AS CR)   Cuby framework   May 21, 2014   8 / 26

Correlated methods

Extremely accurate calculations
post-CCSD(T)

Benchmark calculations
CCSD(T)/CBS
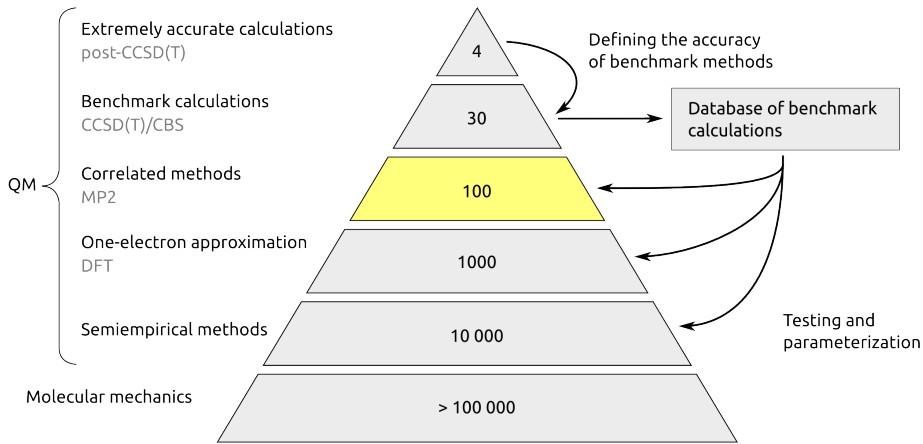
Correlated methods
MP2

One-electron approximation
DFT

Semiempirical methods

Molecular mechanics

QM

4

30

100

1000

10 000

> 100 000

Defining the accuracy
of benchmark methods

Database of benchmark
calculations

Testing and
parameterization

Parameterized correlated methods

- Even advanced correlated methods can bebefit from empirical parameterization
- Near-bechmark results at substantially lower cost: MP2.5[5], SCS-MI-CCSD[6]
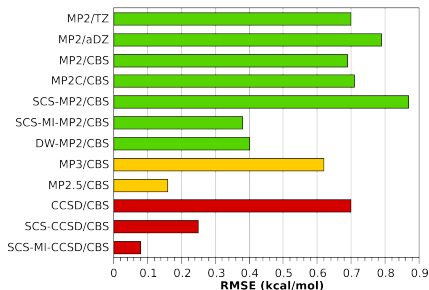
[5]Pitoňák, M.; Neogrády, P.; Černý, J.; Grimme, S.; Hobza, P. ChemPhysChem 2009, 10, 282.

[6]Pitoňák, M.; Řezáč, J.; Hobza, P. Phys. Chem. Chem. Phys. 2010, 12, 9611.

Parameterized correlated methods

- Even advanced correlated methods can bebefit from empirical parameterization
- Near-bechmark results at substantially lower cost: MP2.5[5], SCS-MI-CCSD[6]
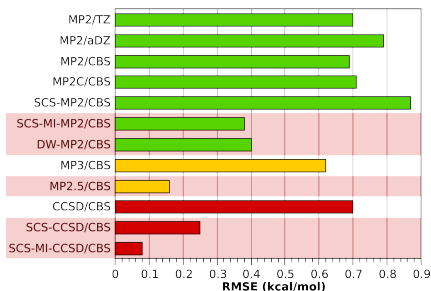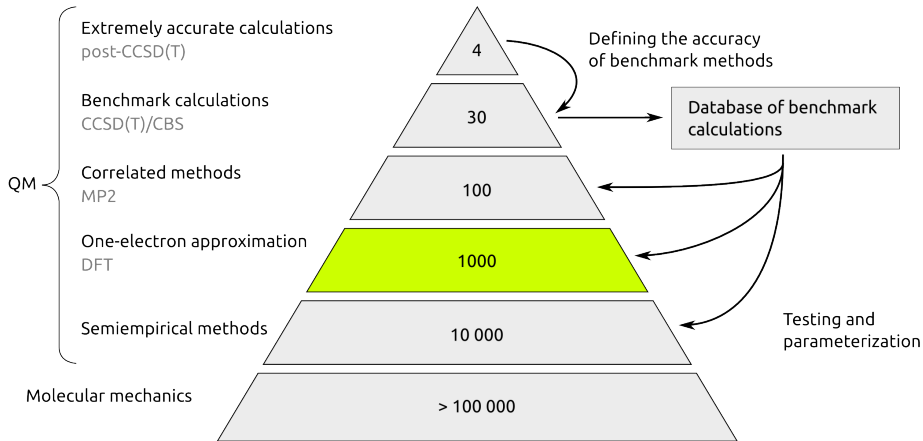

---

[5]Pitoňák, M.; Neogrády, P.; Černý, J.; Grimme, S.; Hobza, P. ChemPhysChem 2009, 10, 282.

[6]Pitoňák, M.; Řezáč, J.; Hobza, P. Phys. Chem. Chem. Phys. 2010, 12, 9611.
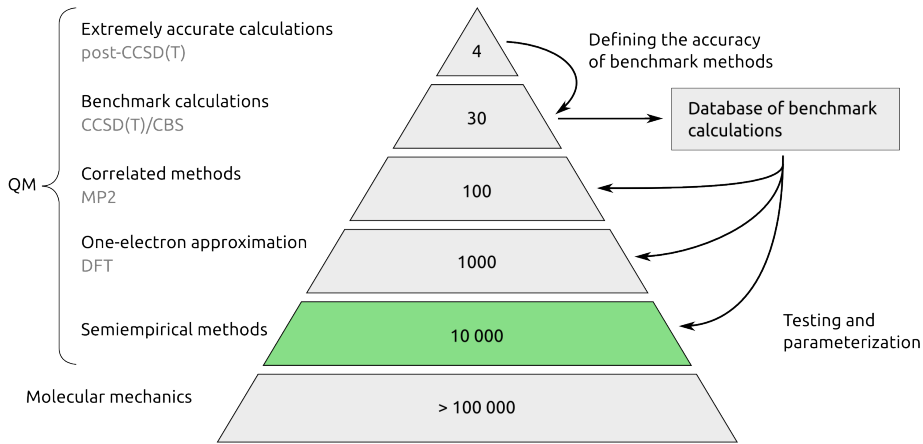
One-electron methods

One-electron methods

DFT-D

Semiempirical methods

Semiempirical methods - Advantages

- Approximate but retain the adavntages of QM calculations
- No system-specific preparation needed
- Very fast, routine calculations of large systems possible[7]
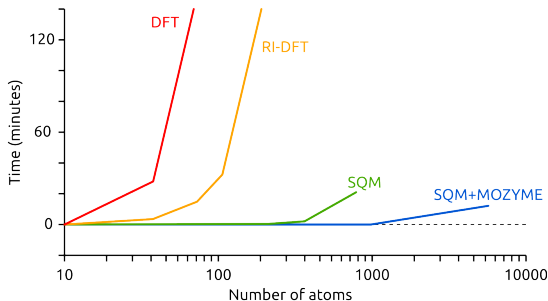
[7]Stewart, J. J. P. J. Mol. Model. 2008, 15, 765.

Semiempirical methods - Problems

- Poor description of noncovalent interactions:
  - London dispersion is missing (one-electron method)
  - Hydrogen bonds underestimated (hydrogen not polarizable)
  - Halogen bonds not described in minimal basis set
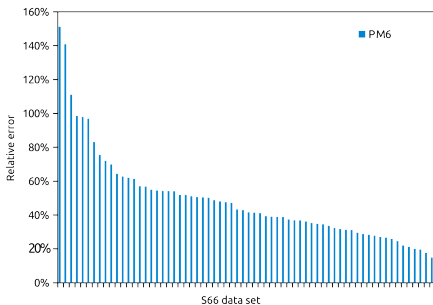
Semiempirical methods - Problems

- Poor description of noncovalent interactions:
  - London dispersion is missing (one-electron method)
  - Hydrogen bonds underestimated (hydrogen not polarizable)
  - Halogen bonds not described in minimal basis set

- Solution: empirical corrections

- Developed for most common semiempirical methods: AM1, PM3, RM1, OM2, PM6

- Applicable also to density functional tight binding (DFTB)

- PM6: Good results, fast, covers almost all elements

- Three generations: PM6-DH[8], PM6-DH2[9], **PM6-D3H4**[10]

---

[8] J. Řezáč, J. Fanfrlík, D. Salahub, and P. Hobza, J. Chem. Theory Comput. 5, 1749 (2009)

[9] M. Korth, M. Pitoňák, J. Řezáč, and P. Hobza, J. Chem. Theory Comput. 6, 344 (2010)

[10] J. Řezáč and P. Hobza, J Chem Theory Comput 8, 141 (2012)

Corrections for semiempirical methods: Dispersion

- Dispersion is missing
- Empirical correction analogous to DFT-D

$$E_{disp} = \sum_i \sum_{j<i} f_{damp}(r_{ij}) \frac{c_{6,ij}}{r_{ij}^6}$$

- PM6-D3H4[11] is based on the advanced D3 correction by S. Grimme[12], uses environment-dependent c_6 coefficients
- Additional correction for underestimated repulsion in hydrocarbons

---

[11] J. Řezáč and P. Hobza, J Chem Theory Comput 8, 141 (2012)

[12] Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. J. Chem. Phys. 132, 154104 (2010)

Corrections for semiempirical methods: Hydrogen bond

- Empirical correction, function of local geometry of the H-bond
- Independent calculation added to the SQM result
- PM6-D3H4 solved all issued of the previous generations

$$E_{HB} = c \cdot f_{rad}(r_{DA}) \cdot f_{ang}(\alpha_{DHA}) \cdot f_{PT}(r_{DH}, r_{HA}) \cdot s_{charged}$$

f_{rad} and f_{ang} are polynomials of desired properties

Corrections for semiempirical methods: Hydrogen bond

- Empirical correction, function of local geometry of the H-bond
- Independent calculation added to the SQM result
- PM6-D3H4 solved all issued of the previous generations

	H2	H+	H4
exact gradient	NO	YES	YES
proton transfer	NO	YES	YES
accurate for charged systems	YES	NO	YES
smooth energy derivatives	NO	NO	YES
coordinates per bond (torsions)	4 (2)	7 (4)	3 (0)

# Corrections for semiempirical methods: Halogen bond

- -X correction[13], specific for PM6 where repulsion was strongly underestimated
- Later reparameterized on better benchmarks[14]
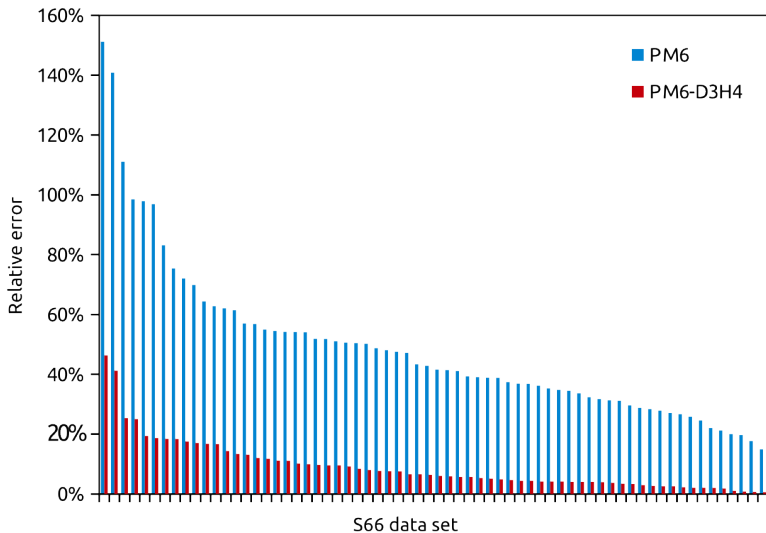- PM6-D3H4X can describe wide range of interactions with chemical accuracy

---
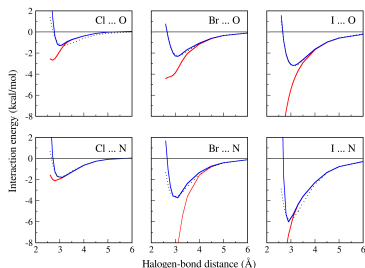
[13] J. Řezáč and P. Hobza, Chem. Phys. Lett. 506, 286 (2011)

[14] Brahmkshatriya, P. S. et al., Curr. Comput.-Aid. Drug. 2013, 9, 118–129.

Results - S66 data set
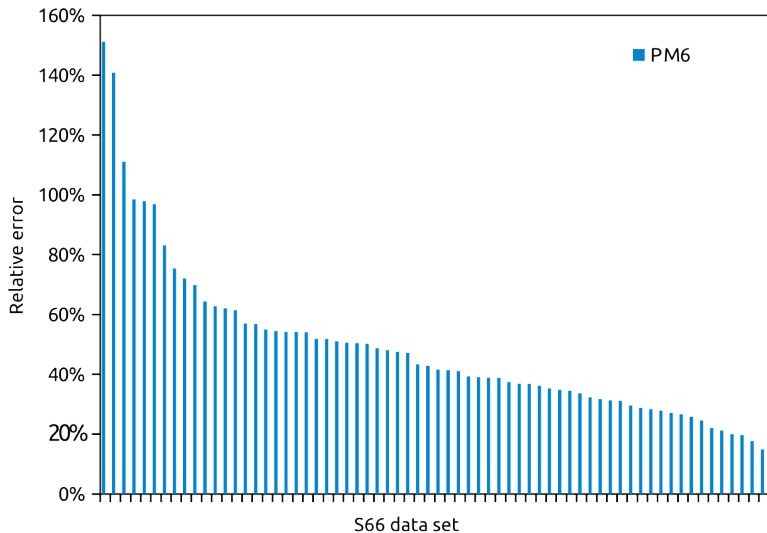
Results - S66 data set

Results - comparison

	S66	S66x8	S66a8	S22	H-bonds	Charged HB	Hydrocarbons	AASidechains	AVG	AVG$_{testing}$
PM6	3.02	2.49	2.12	4.16	3.18	3.92	2.64	4.08	3.20	3.30
PM6-DH2	0.91	0.79	0.73	0.54	1.52	2.21	0.67	1.32	1.09	1.17
PM6-DH+	0.82	0.76	0.67	0.80	1.43	1.94	0.67	1.89	1.12	1.33
PM6-D3H4	0.65	0.66	0.68	0.78	1.05	1.11	0.71	1.17	0.85	0.98
PM6-D3H4*	0.70	0.71	0.74	0.84	1.12	2.26	0.71	1.86	1.12	1.23
DFTB	2.88	2.40	2.24	3.45	2.82	4.78	2.90	3.44	3.11	3.05
DFTB-D	1.50	1.43	1.28	1.63	1.96	4.28	0.59	2.27	1.87	1.60
DFTB-D,γ	1.17	1.17	1.04	1.21	1.61	3.67	0.56	1.82	1.53	1.33
DFTB-DH2	1.44	1.15	0.98	1.86	1.54	2.13	0.59	1.62	1.41	1.25
DFTB-D3H4	0.67	0.62	0.61	0.97	0.71	1.43	0.59	0.88	0.81	0.73
RM1	5.39	4.38	4.13	7.15	5.40	5.60	3.65	5.34	5.13	4.80
RM1-D3H4	0.92	0.90	0.78	1.03	0.90	2.05	0.24	0.73	0.94	0.62
RM1-D3H4*	0.91	0.90	0.79	1.03	0.89	2.09	0.24	0.93	0.97	0.69
OM3[a]	3.33	2.70	2.49	4.17	2.88	3.00	3.93	4.99	3.44	3.93
OM3-DH2[a]	0.80	0.96	0.62	0.96	0.84	1.83	1.11	1.53	1.08	1.16
OM3-D3H4[a]	0.48	0.60	0.42	0.58	0.56	1.50	0.70	2.34	0.90	1.20
AM1	6.24	5.27	4.03	8.66	6.10	7.64	3.73	6.38	6.01	5.40
AM1-DH2	1.93	1.96	1.47	0.85	2.08	3.58	3.94	3.71	2.44	3.25
AM1-D3H4	1.35	1.76	1.45	1.76	2.11	3.04	0.82	2.02	1.79	1.65
PM3	5.08	4.51	3.77	7.64	4.98	7.03	2.25	4.60	4.98	3.94
PM3-D3H4	1.40	1.26	0.97	2.51	0.83	2.23	0.40	1.05	1.33	0.76
TPSS/TZVP-D	0.69	0.53	0.57	0.58	1.04	1.89	0.72	0.89	0.86	0.88
MP2/cc-pVTZ	0.70	0.59	0.57	1.85	1.40	1.81	0.88	1.62	1.18	1.30

Errors in kcal/mol

Applications of the methods

- In silico drug design[15,16]
  - Optimization of whole protein or large part of it
  - Reliable protein-ligand interaction energies
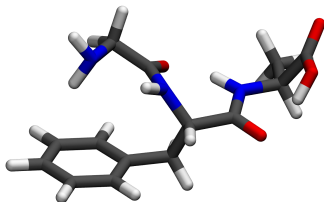 - Can reproduce and predict experiment
  - Outperforms MM forcefield
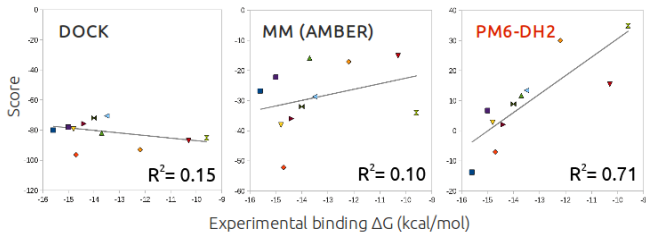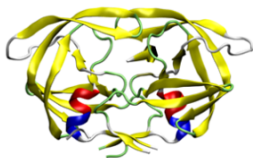
[15]Fanfrlík, J. et al. Phys. Chem. B 2010, 114, 12666–12678.

[16]Lepšík, M.; Řezáč, J.; Kolář, M.; Pecina, A.; Hobza, P.; Fanfrlík, J. ChemPlusChem 2013

Peptide conformations

- GFA tripeptide, CCSD(T) benchmark for 16 low-lying conformers[17]
- Relative energy range 1.6 kcal/mol

---

[17]Valdes, H.; Pluháčková, K.; Pitoňák, M.; Řezáč, J.; Hobza, P. Phys. Chem. Chem. Phys. 2008, 10, 2747.

Peptide conformations

- GFA tripeptide, CCSD(T) benchmark for 16 low-lying conformers[17]
- Relative energy range 1.6 kcal/mol

Method	RMSE (kcal/mol)
AMBER ff, water charges	2.4
AMBER ff, gas phase charges	1.4
MP2/CBS	0.4
BLYP-D3	0.8
DFTB-D	0.6
DFTB-D3	0.4
DFTB-D3H4	1.1
PM6	1.3
PM6-D3H4	0.7

---

[17]Valdes, H.; Pluháčková, K.; Pitonák, M.; Řezáč, J.; Hobza, P. Phys. Chem. Chem. Phys. 2008, 10, 2747.

PM7

- The latest semeiempirical method PM7[18] already contains corrections derived from PM6-DH2
- It performs similarly to PM6-D3H4 in small complexes[19]
- It strongly overetsimates the interactions in larger systems, we are working on a fix

---

[18]Stewart, J. J. P. J Mol Model 2013, 19, 1–32.

[19]J. Hostaš, J. Řezáč, and P. Hobza, Chem. Phys. Lett. 568–569, 161 (2013)

Data and method availability

- Our benchmark data sets are available online at www.begdb.com
- Corrected SQM methods up to PM6-DH2 are available in MOPAC
- The -D3H4 correction implementation is available at www.molecular.cz/~rezac

Summary

- Large database of benchmark data is indispensable for development of parameterized methods

- It is important to use also nonequilibrium geometries

- Parameterized correlated methods such as MP2.5 and SCS-MI-CCSD are economic alternative to CCSD(T) calculations

- Corrected SQM methods can describe noncovalent interactions with error < 1 kcal/mol

- Applicable to thousands of atoms

Acknowledgements

- Pavel Hobza
- Kevin Riley (USA)
- Robert Sedlák

- Michal Pitoňák (Slovakia)

- Jindřich Fanfrlík
- Martin Lepšík

- Haydee Valdes (Spain)