Computational chemistry today, or how to obtain accurate numbers with minimum effort

Zlatko Mihalić

mihalic@chem.pmf.hr

University of Zagreb, Faculty of Science, Department of Chemistry
Horvatovac 102a, 10000 Zagreb, Croatia

SOME OF THE KEY MOMENTS IN EARLY HISTORY OF COMPUTATIONAL CHEMISTRY

1926	Schrödinger equation (WFT)
1927	Heitler and London: first calculations in theoretical chemistry
1927	Thomas–Fermi method (first DFT)
1935	Pauling and Wilson: *Introduction to Quantum Mechanics – with Applications to Chemistry*
1944	Eyring, Walter, Kimball: *Quantum Chemistry*
1952	Coulson: *Valence*
1964	Hohenberg–Kohn theorem(s)
1965	Kohn–Sham equations: the beginning of modern DFT
1965	Wiberg: first general molecular mechanics (MM) program
1970	Pople (Carnegie Mellon University): Gaussian 70 (HF calculations)
1976	Allinger: MM1 and (in 1977) MM2 force fields
1980	*Journal of Computational Chemistry*
1987	Gaussian Inc.
1990s	First modern functionals (GGA, mGGA, hybrid functionals)
1992	Gaussian 1992/DFT

FIRST NOBEL PRIZE FOR COMPUTATIONAL CHEMISTRY (1998)
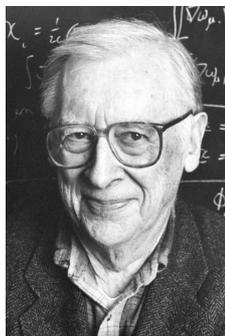
Walter Kohn   John A. Pople

The Nobel Prize in Chemistry 1998 was divided equally between Walter Kohn "for his development of the density-functional theory" and John A. Pople "for his development of computational methods in quantum chemistry."

SECOND NOBEL PRIZE FOR COMPUTATIONAL CHEMISTRY (2013)

Martin Karplus   Michael Levitt   Arieh Warshel

The Nobel Prize in Chemistry 2013 was awarded jointly to Martin Karplus, Michael Levitt and Arieh Warshel "for the development of multiscale models for complex chemical systems".

LEVEL OF THEORY (MODEL CHEMISTRY)

Ab initio (WFT), based on wave function (generally complex function of $4N$ variables)

- Hartree-Fock (HF) – wave function is (Slater) determinant composed of spin-orbitals

- post-Hartree-Fock (MPn, CI, CC) – wf is linear combination of Slater determinants

Density functional theory (DFT), based on electron density (real function of 3 variables)

In both cases, one-electron functions (orbitals) are used to facilitate the calculations. Orbitals are approximated by linear combination of simpler functions, elements basis sets. If basis set is complete (infinte number of basis functions), the relationship is exact.

Since we use noninfinite basis sets, various errors occur: BSSE, BSIE, ...

Pople's designation of level of theory (model chemistry) consists of method and basis set:

METHOD/BASIS SET

Often higher level of theory is used for the energy calculation following the geometry optimization:

METHOD-2/BASIS SET-2//METHOD-1/BASIS SET-1

Popular Theoretical Methods Predict Benzene and Arenes To Be Nonplanar

Damian Moran,[*,†,‡] Andrew C. Simmonett,[†] Franklin E. Leach III,[†] Wesley D. Allen,[*,†]
Paul v. R. Schleyer,[*,†] and Henry F. Schaefer III[†]

*Center for Computational Chemistry, University of Georgia, Athens, Georgia 30602, and
School of Chemistry, University of Sydney, Sydney, NSW 2006, Australia*

Received May 1, 2006; E-mail: dmoran@chem.usyd.edu.au; wdallen@ccc.uga.edu; schleyer@chem.uga.edu

Remarkably, ab initio computations on benzene (**1**) at electron-correlated MP2, MP3, CISD, and CCSD levels using a number of popular basis sets[1] give anomalous, nonplanar equilibrium structures. For example, planar (D_{6h}) benzene has at least one sizable imaginary vibrational frequency at MP2/6-311G ($723i$ cm^{-1}, b_{2g}), MP2/6-311++G ($1844i$ cm^{-1}, b_{2g}; $462i$ cm^{-1}, e_{2u}), MP3/6-31++G-(d,p) ($862i$ cm^{-1}, b_{2g}), CISD/6-311G ($190i$ cm^{-1}, b_{2g}), and CISD/6-311++G(d,p) ($511i$ cm^{-1}, b_{2g}). Similar unexpected results at the same levels of theory are found for other planar aromatic molecules: naphthalene (**2**), the cyclopentadienyl (**3**) and indenyl anions (**4**), the tropylium cation (**5**), anthracene (**6**), and pyridine (**7**) (Figure 1). In contrast, RHF, BLYP, and B3LYP computations[1] with the same basis sets exclusively yield real frequencies for **1**−**7**. Extensive frequency tabulations at numerous levels of theory are provided as Supporting Information. These anomalous results serve as stark warnings for black-box ab initio studies of aromatic hydrocarbons, particularly large, polycyclic species for which only limited basis sets are feasible.

1 (D_{6h}) **2** (D_{2h}) **3** (D_{5h}) **4** (C_{2v})

5 (D_{7h}) **6** (D_{2h}) **7** (C_{2v})
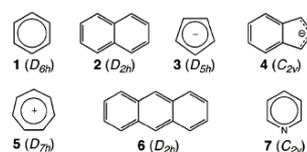
Figure 1. Arene nonplanarity failures of popular ab initio methods.

LEVEL OF THEORY ADEQUATE FOR NONCOVALENT INTERACTIONS
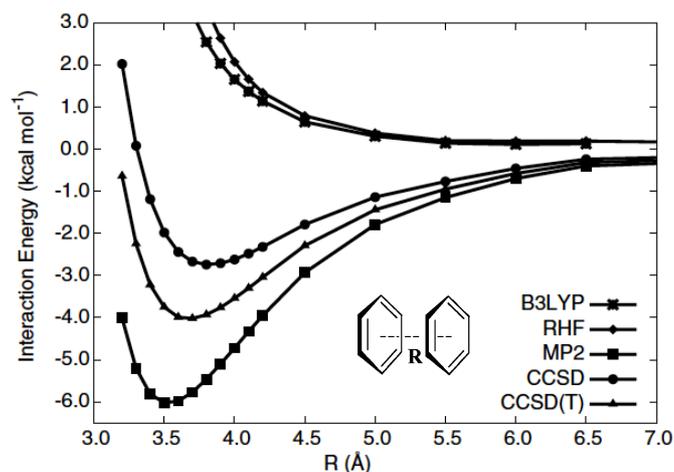
Interaction energy for the sandwich benzene dimer as a function of the intermolecular distance using various methods and the aug-cc-pVDZ basis, with the rigid monomer geometries. Energies not cp corrected.
(C. David Sherrill, "Computations of Noncovalent π Interactions", *Rev. Comp. Chem.* 2008, Volume 26, 1–30.)

SPEEDING UP MP2 CALCULATIONS

Second-order Møller-Plesset perturbation theory (MP2) is the computationally least expensive and therefore the most popular *ab initio* electron correlation method. It typically recovers 80-90% of the correlation energy, and scales as N^5 with the system size. There are several approaches to how to speed up once prohibitively slow MP2 calculations.

Resolution-of-identity approximation (RI-MP2)

Local correlation approximation (LMP2, LPNO-MP2, DLPNO-MP2)

Spin-component-scaled (SCS-MP2) and spin-opposite-scaled (SOS-MP2) approximations

Explicitly correlated R12 and F12 approaches (MP2-R12/F12)

RESOLUTION-OF-IDENTITY (RI) APPROXIMATION

For a molecule of fixed size, MP2 computational cost is dominated by the integral transformation from the atomic to the molecular orbital basis. In resolution-of-identity approximation (RI) (or density fitting, DF), this four-index transformation is circumvented by the integral approximation which reduces the required time without significant loss in accuracy. Using auxilary basis set, P, method approximates basis functions products $\nu\mu$ as

$$\nu(\mathbf{r})\mu(\mathbf{r}) \approx \sum_i c_{\nu\mu}^P P_i(\mathbf{r})$$

so that the four-index electron repulsion integrals

$$(\nu\mu|\rho\sigma) = \int \mu(\mathbf{r}_1)\nu(\mathbf{r}_1)\frac{1}{\mathbf{r}_{12}}\rho(\mathbf{r}_2)\sigma(\mathbf{r}_2)\,d\mathbf{r}_1 d\mathbf{r}_2$$

can be approximated by simpler two- and three-index integrals.

For not too large molecules, the RI-MP2 treatment of electron correlation is so efficient that it lasts not much longer than the initial Hartree-Fock calculation, which, by the way, can also be sped up by the resolution of identity approximation (RI-JK).

RESOLUTION-OF-IDENTITY (RI) APPROXIMATION
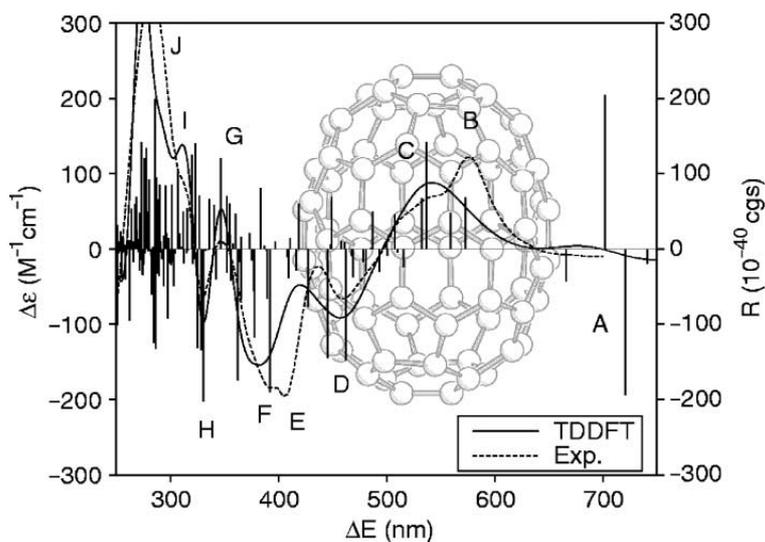
Calculated and experimental CD spectra of fulerene C_{76} (BP86/SVP i RI/J approximation).

LOCAL CORRELATION APPROXIMATION

Another way to speed up MP2 computations is to apply the local approximation. As electron repulsion between two products in the integral $(\mu\nu|\rho\sigma)$ depends inversely on the distance $\mathbf{r}_{12}$ between two electrons, at large distances this interaction can be approximated or neglected. After localizing orbitals, LMP2 method considers only interactions of electrons in close orbitals.

Edmiston and Krauss introduced the concept of PNOs, where the basic idea consists of using set of natural orbitals (NOs) to correlate each electron pair. This triggered the development of correlation methods based on PNOs.

Program ORCA has two local methods, LPNO ("local pair natural orbital") and DLPNO ("domain based LPNO"), designed to provide 'results as close as possible to the canonical coupled-cluster results while gaining orders of magnitude of efficiency' (ORCA 4.1 manual). The DLPNO has linear scaling and is of similar accuracy as older LPNO method.

LOCAL CORRELATION APPROXIMATION – DLPNO-CCSD(T) METHOD

D. G. Liakos and F. Neese "Is It Possible To Obtain Coupled Cluster Quality Energies at near Density Functional Theory Cost? Domain-Based Local Pair Natural Orbital Coupled Cluster vs Modern Density Functional Theory", *J. Chem. Theory Comput.,* 2015, 11, 4054–4063.

B3LYP $3.58e-04 * N^{2.26}$
NormalPNO $1.43e-03 * N^{2.11}$
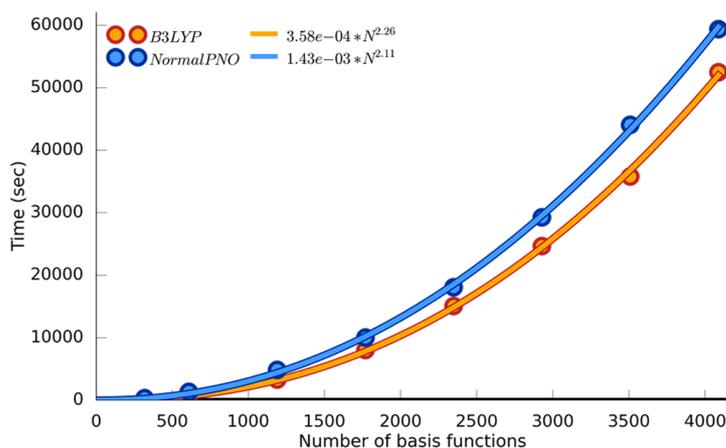
Scaling of the NormalPNO DLPNO-CCSD(T) and B3LYP methods in conjunction with cc-pVTZ for linear alkanes C_nH_{2n+2} for n up to 70.

SCS-MP2 AND SOS-MP2 APPROXIMATIONS

Grimme (2003) introduced different scale factors for the same-spin (E_T) and opposite-spin pair (E_S) correlation energies, yielding the spin-component-scaled (SCS) MP2 method.

$$E_c \approx E_c^{\text{SCS-MP2}} = p_S E_S^{(2)} + p_T E_T^{(2)}$$

Related to the SCS method is Head-Gordon's spin-opposite-scaled MP2 (SOS-MP2) method, which completely neglects the same-spin term and scales up the opposite-spin term. SOS-MP2 is order of magnitude faster than MP2 and SCS-MP2 methods.
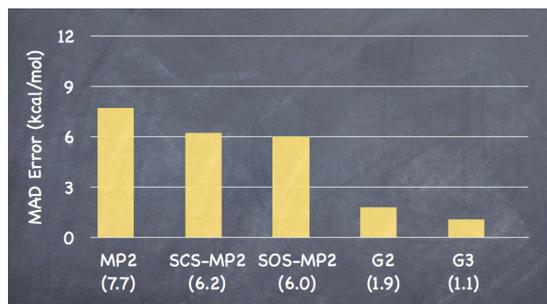
MAD errors for bond dissociation energies (92 energies)
(from Q-Chem web site: Y. Jung, 'MP2 and its Variants in Q-Chem')

EXPLICITLY CORRELATED R12 AND F12 METHODS

One way to accelerate higher correlation methods basis set convergence towards CBS limit it is to employ explicitly correlated methods, in which the electronic wave function depend explicitly on the distances between electrons (Hylleeras, 1929). The first such successful approach is Kutzelnigg's "linear R12" method (1985), in which alongside products of one-electron functions

$$\nu(\mathbf{r}_1)\mu(\mathbf{r}_2)$$

terms linear in interelectronic distance are also used

$$\mathbf{r}_{12}\nu(\mathbf{r}_1)\mu(\mathbf{r}_2)$$

In newer F12 methods, faster convergence to the CBS limit is achieved by replacing the linear R12 terms with more complex functions of the interelectronic distance

$$f(\mathbf{r}_{12})\nu(\mathbf{r}_1)\mu(\mathbf{r}_2)$$

Today a number variants of F12 approach exist, implemented in various quantum chemistry programs.

EXPLICITLY CORRELATED R12 AND F12 METHODS
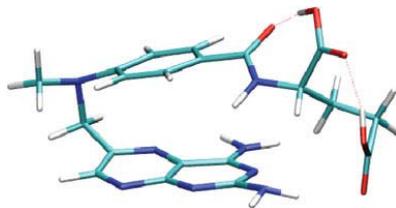
Basis	Size	ΔE(MP2)	%	ΔE(MP2-F12)	%
aug-cc-pVDZ	957	-5.157	76.3	-6.644	98.4
aug-cc-pVTZ	2024	-6.142	90.9	-6.718	99.5
aug-cc-pVQZ	3652	-6.480	95.9	-6.745	99.9
aug-cc-pV5Z	5951	-6.611	97.9		
CBS limit		**-6.755**			
cc-pVDZ-F12	1188	-5.604	83.0	-6.690	99.0
cc-pVTZ-F12	2145	-6.257	92.6	-6.714	99.4
cc-pVQZ-F12	3619	-6.510	96.4	-6.755	100.0

Correlation energy of methotrexane optimized at MP2/aug-cc-pVDZ level of theory

(Bachorz et al., "The MP2-F12 Method in the Turbomole Program Package", *J. Comp. Chem.,* 2011*,* 32(11), 2492.)

DENSITY FUNCTIONAL THEORY (DFT)

Kohn and Hohenberg proved that the energy of molecular system is uniquelly determined by its electron density. Corresponding energy functional has three components: kinetic energy of electrons, nuclear-electron attraction and electron-electron repulsion.

$$E[\rho] = T[\rho] + V_{ne}[\rho] + V_{ee}[\rho]$$

Kohn and Sham considered a fictious system consisting od non-interactiong electrons (represented by KS orbitals, φ_i), which has the same electron density as the real system

$$\rho(\mathbf{r}) = \sum_i^N |\varphi_i(\mathbf{r})|^2$$

For such a system, energy functional can be written as sum of three easily calculated terms, and unknown exchange-correlation term, E_{xc}

$$E_{DFT}[\rho] = T_{ni}[\rho] + E_{ne}[\rho] + E_{ee}[\rho] + E_{xc}[\rho]$$

which is usually considered as a sum of two parts, exchange and correlation functionals.

$$E_{xc}[\rho] = E_x[\rho] + E_c[\rho]$$

A hughe number of today existing approximations of exchange-correlation functional (DFAs) poses serious dilemma to non-specialist users.

JACOB'S LADDER OF DFAs

(Picture from the paper: L. Goerigk, N. Mehta, "A Trip to the Density Functional Theory Zoo: Warnings and Recommendations for the User", *Aust. J. Chem.* 2019)

HEAVEN OF CHEMICAL ACCURACY

GGA/meta-GGA + **HF** + **MP2** — DOUBLE-HYBRIDS (DHDFs)

GGA/meta-GGA + **HF** — HYBRIDS

$\sum_i \left| \nabla \varphi_i(\mathbf{r}) \right|^2$ or $\nabla^2 \rho(\mathbf{r})$ — meta-GGA

$\nabla \rho(\mathbf{r})$ — GENERALIZED GRADIENT APPROXIMATION (GGA)

$\rho(\mathbf{r})$ — LOCAL DENSITY APPROXIMATION

HARTREE HELL
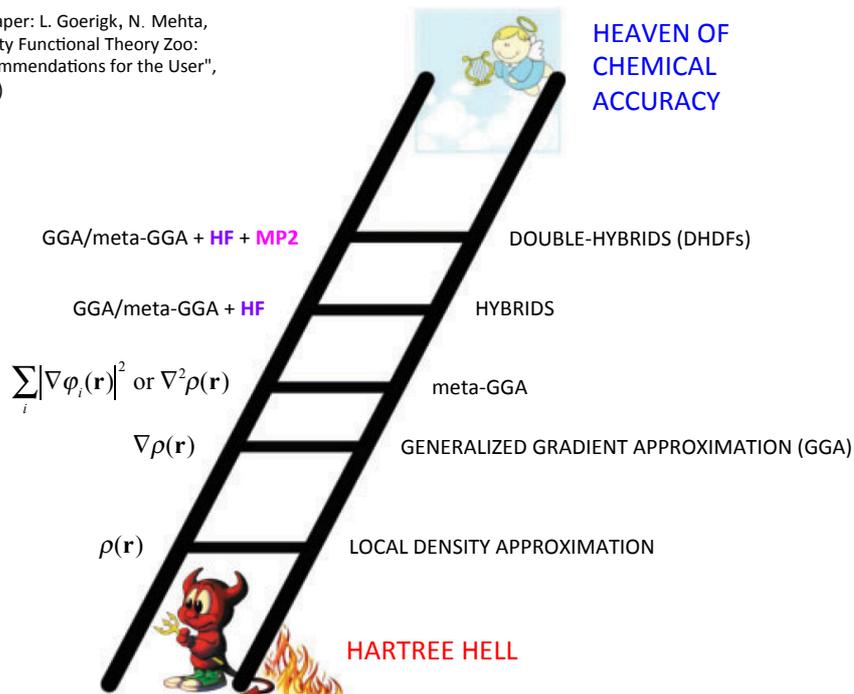
Exchange-Correlation Functionals (Q-Chem 5.1 User's Manual)

SPW92 LDA SVWN5 B97-D3(0) B97-D PBE BLYP revPBE BEEF-vdW BOP BP86 BP86VWN BPBE EDF1 EDF2 GAM HCTH93 HCTH120 HCTH147 HCTH407 HLE16 KT1 KT2 KT3 mPW91 N12 OLYP PBEOP PBEsol PW91 RPBE rVV10 SOGGA SOGGA11 VV10 B97M-V B97M-rV M06-L TPSS revTPSS BLOC M11-L mBEEF MGGA_MS0 MGGA_MS1 MGGA_MS2 MGGA_MVS MN12-L MN15-L oTPSS PKZB SCAN τ-HCTH TM VSXC B3LYP PBE0 revPBE0 B97 B1LYP B1PW91 B3LYP5 B3P86 B3PW91 B5050LYP B97-1 B97-2 B97-3 B97-K HFLYP MPW1K MPW1LYP MPW1PBE MPW1PW91 O3LYP PBEh-3c PBE50 SOGGA11-X WC04 WP04 X3LYP M06-2X M08-HX TPSSh revTPSSh B1B95 B3TLAP BB1K BMK dlDF M05 M05-2X M06 M06-HF M08-SO MGGA_MS2h MGGA_MVSh MN15 MPW1B95 MPWB1K PW6B95 B95 PWB6K SCAN0 τ-HCTHh TPSS0 ωB97X-V ωB97X-D3 ωB97X-D CAM-B3LYP CAM-QTP00 CAM-QTP01 HSE-HJS LC-rVV10 LC-VV10 LC-ωPBE08 LRC-BOP LRC-ωPBE LRC-ωPBEh N12-SX rCAM-B3LYP ωB97 ωB97X ωB97X-rV ωB97M-V M11 ωB97M-rV ωM05-D ωM06-D3 DSD-PBEPBE-D3 ωB97X-2(LP) ωB97X-2(TQZ) XYG3 XYGJ-OS B2PLYP B2GPPLYP DSD-PBEP86-D3 LS1DH-PBE PBE-QIDH PBE0-2 PBE0-DH PTPSS-D3 DSD-PBEB95-D3 PWPB95-D3 SRC1-R1 SRC1-R2 SRC2-R1 SRC2-R2 BR89 B94 B94hyb BR89B94h BRSC MB05 B05 BM05(XC) PSTS MCY2

Marcel Swart's DFT2018 Pool

First Division	Second Division
B2PLYP	B3PW91
B3LYP	BHandH
B3LYP-D	BLYP
B97-D	DSD-BLYP
B97M-V	DSD-PBEP86
BP86	LC-PBE
CAM-B3LYP	LDA
HSE	M06
LC-ωPBE	M06-L
M06-2X	OLYP
PBE	optB88-vdW
PBE-D	PW6B95
PBE0	revTPSS
PW91	revTPSS-D
PWPB96-D3	RPA
revPBE	RPBE
TPSSh	S12g
ωB97M-V	SAOP
ωB97X-D	SCAN
ωB97X-V	SSB-D

Still used today! → B3LYP

IS B3LYP/6-31G* REALLY COST-EFFECTIVE LEVEL OF THEORY?

Why the Standard B3LYP/6-31G* Model Chemistry Should Not Be Used in DFT Calculations of Molecular Thermochemistry: Understanding and Correcting the Problem

Holger Kruse, Lars Goerigk, and Stefan Grimme, *J. Org. Chem.* **2012**, 77 (23), 10824–10834.

B3LYP/6-31G*   B3LYP-gCP-D3/6-31G*

Dispersion

BSSE

BSSE   Dispersion
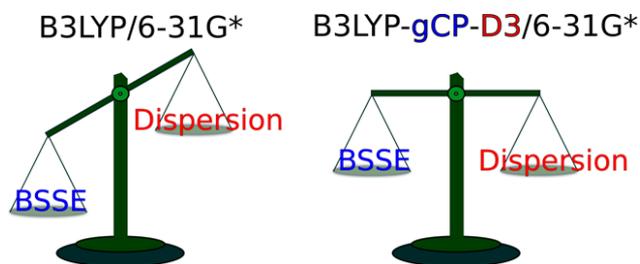
Goal: To get (at least) qualitatively correct numbers fast, but for correct reasons!

DATABASES FOR STATISTICAL EVALUATION (BENCHMARKING) OF DFAs

GMTKN55
55 sets
2462 single point calculations
1505 relative energies

barrier heights
7 sets
312 single point calculations
194 relative energies

reaction energies for
large systems and
isomerisation reactions
9 sets
394 single point calculations
243 relative energies

basic properties
and reaction energies for
small systems
18 sets
718 single point calculations
473 relative energies

intermolecular
noncovalent interactions
12 sets
699 single point calculations
304 relative energies

intramolecular
noncovalent interactions
9 sets
339 single point calculations
291 relative energies

noncovalent interactions
21 sets
1038 single point calculations
595 relative energies

(**G**ENERAL **M**AIN GROUP **T**HERMOCHEMISTRTY, **K**INETICS AND **N**ONCOVALENT INTERACTIONS)

DATABASES FOR STATISTICAL EVALUATION (BENCHMARKING) OF DF(A)s

A number of such databases exists: GMTKN, MGCDB84, Minnesota2015B, DP284, W4-17, ACCDB, to name the most important ones. They offer objective base for validation of new DFAs, as well as for comparison of the existing ones.

Their importance can be illustrated by the fact that Google finds more than 1300 hits for the oldest version of GMTKN database, GMTKN24, 7000 hits for GMTKN30 and 1700 hits for GMTKN55.

Three recent papers that objectively compare the most important DFAs:

N. Mardirossian, M. Head-Gordon, "Thirty years of density functional theory in computational chemistry: an overview and extensive assessment of 200 density functionals", *Molecular Physics*, 2017, 115(19), 2315-2372.

L. Goerigk, A. Hansen, C. Bauer, S. Ehrlich, A. Najibi, S. Grimme, "A look at the density functional theory zoo with the advanced GMTKN55 database for general main group thermochemistry, kinetics and noncovalent interactions", *Phys. Chem. Chem. Phys.*, 2017, 19, 32147–32744.

L. Goerigk, N. Mehta, "A trip to the density functional theory zoo: Warnings and recommendations for the user", *Aust. J. Chem.* 2019, https://doi.org/10.1071/CH19023

WATER CLUSTERS

Obtaining reliable theoretical descriptions of the simplest water clusters is very important for understanding the solvation effects and the complex hydrogen-bonding dynamics in liquid water.

The water dimer has long served as a model for hydrogen-bonding interactions. The global minima for clusters with less that five molecules are characterized by cyclic homodromic hydrogen-bonding networks, *i.e.* each water molecule accepts and donates one hydrogen bond.

Larger clusters are different, offering insight into hydrogen-bonding cooperativity.

How then to calculate benchmark-quality *ab initio* energies of clusters with six or more molecules?

First, one has to find level of theory that produces good geometries. The obvious choice is to use DFT, since geometry optimization of larger clusters on higher correlation levels is hardly an option.

The qualities of geometries obtained at different DFT levels of theory can be judged by comparison with the data from the WATER27 database (part of GMTKN55).

WATER CLUSTERS
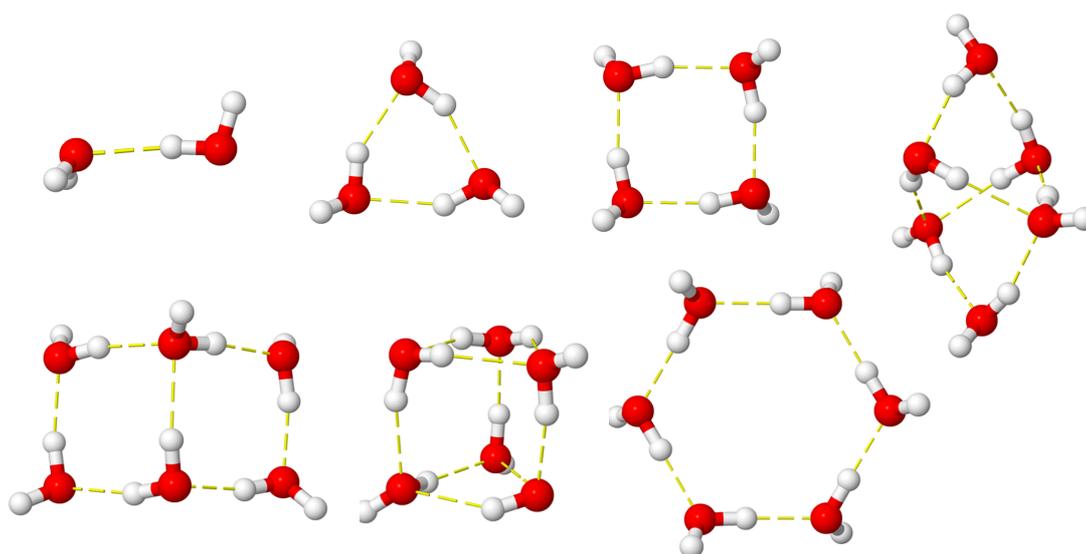
Geometries of water clusters $(H_2O)_n$ (n = 1–6) were calculated at various DFT levels of theory and compared to corresponding geometries from WATER27 database.

Time required for different QM programs to optimize the same starting geometry of cyclic $(H_2O)_6$ cluster.
Double-hybrid B2PLYP-D3 (blue) with RI- approximation can be faster than B3LYP-D3 (pink) in other program!

Average RMSDs between benchmark WATER27 geometries and geometries of eight
shown water clusters calculated at DF/aug-cc-pVTZ level of theory.
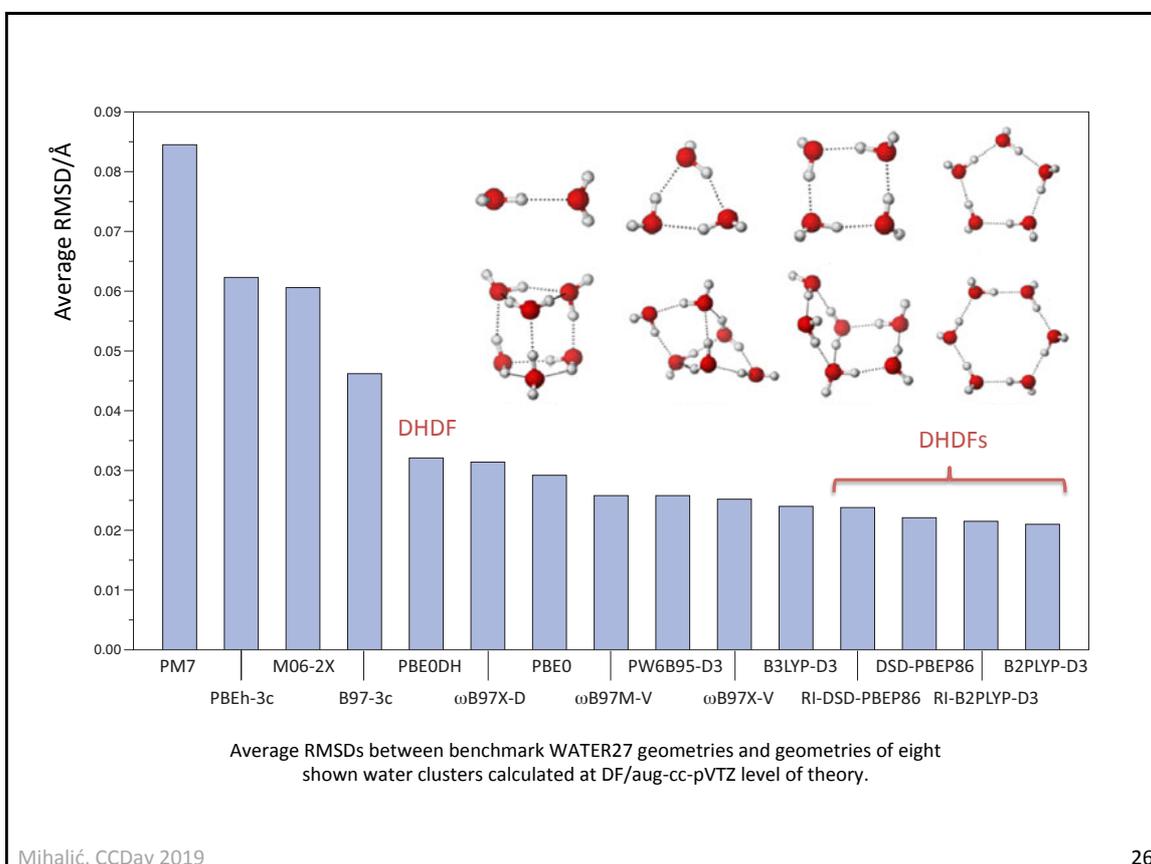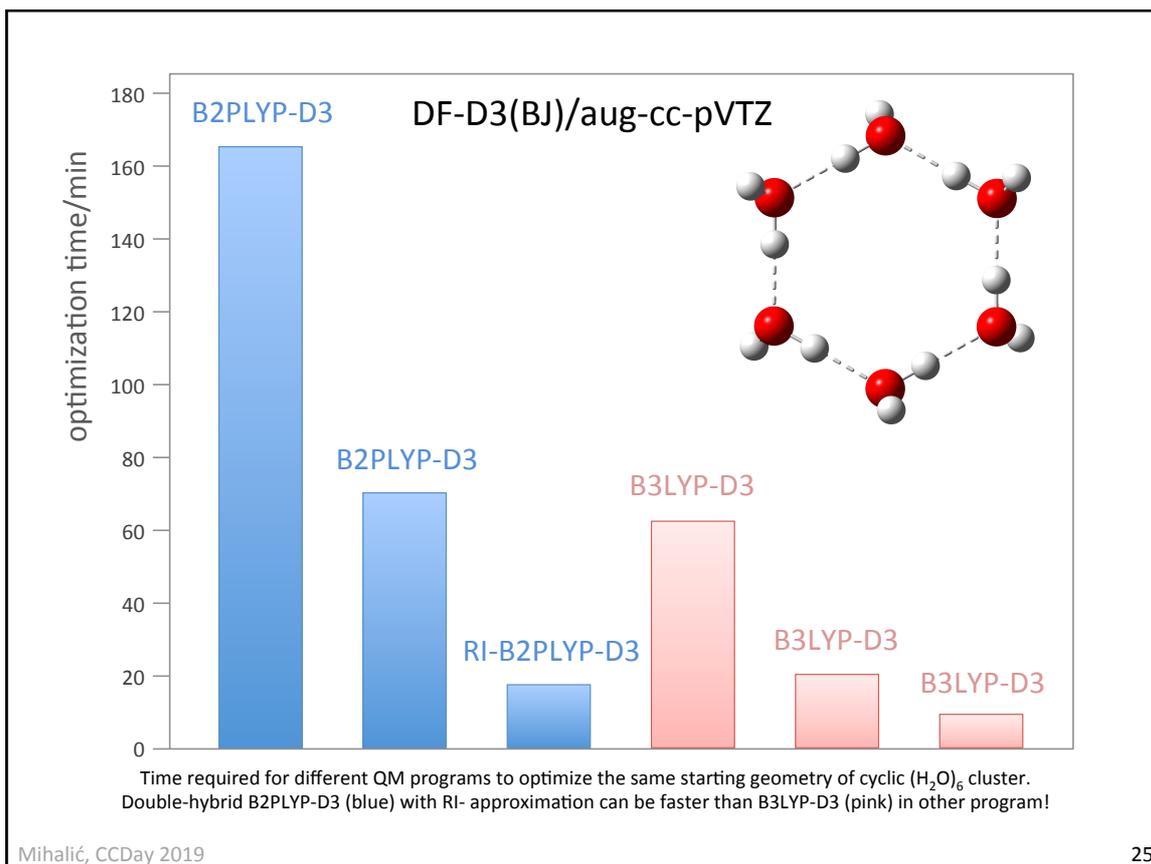
MONOMER–DIMER EQUILIBRIA OF AROMATIC *C*-NITROSO COMPOUNDS
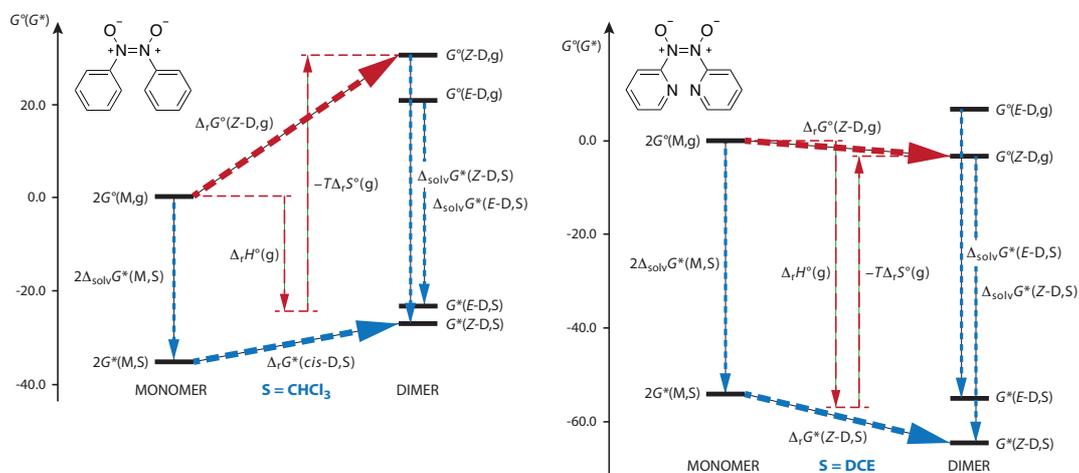
K. Varga, I. Biljan, V. Tomišić, Z. Mihalić, H Vančik, "Quantum Chemical Calculations of Monomer–Dimer Equilibria of Aromatic C-Nitroso Compounds ", *J. Phys. Chem A*. 2018, 122(9), 2542-2549.

MONOMER–DIMER EQUILIBRIA OF AROMATIC *C*-NITROSO COMPOUNDS
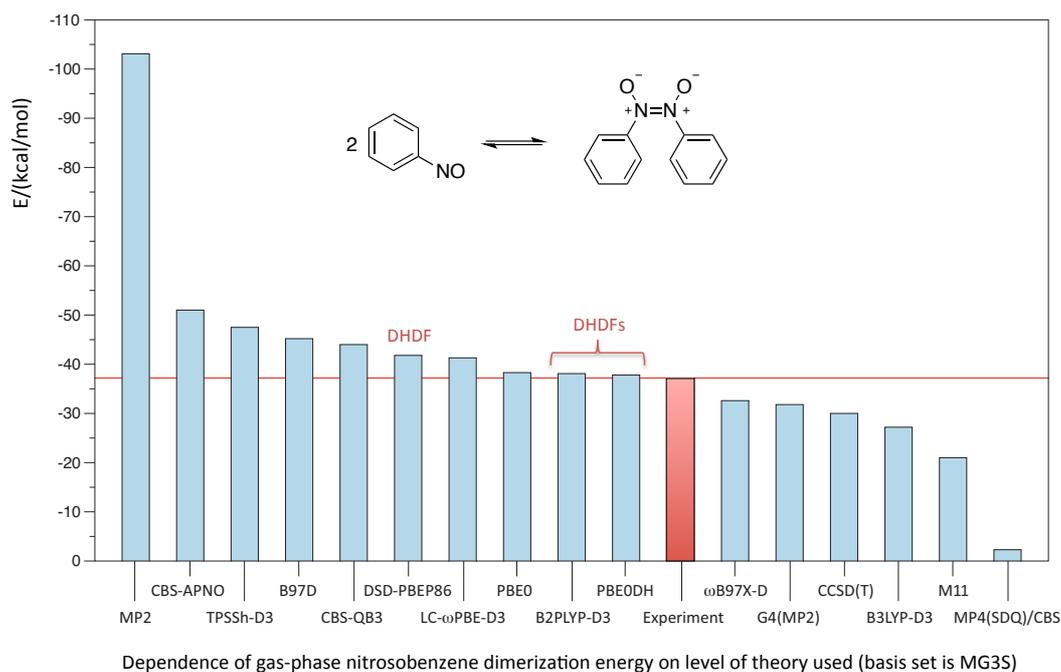
Dependence of gas-phase nitrosobenzene dimerization energy on level of theory used (basis set is MG3S)