

A EUROPEAN JOURNAL

CHEMPHYSICHEM

OF CHEMICAL PHYSICS AND PHYSICAL CHEMISTRY

Accepted Article

Title: A Robust and Cost-Efficient Scheme for Accurate Conformational Energies of Organic Molecules

Authors: Luigi Cavallo, Yury Minenkov, Dmitry Sharapa, and Alexander Genaev

This manuscript has been accepted after peer review and appears as an Accepted Article online prior to editing, proofing, and formal publication of the final Version of Record (VoR). This work is currently citable by using the Digital Object Identifier (DOI) given below. The VoR will be published online in Early View as soon as possible and may be different to this Accepted Article as a result of editing. Readers should obtain the VoR from the journal website shown below when it is published to ensure accuracy of information. The authors are responsible for the content of this Accepted Article.

To be cited as: *ChemPhysChem* 10.1002/cphc.201801063

Link to VoR: <http://dx.doi.org/10.1002/cphc.201801063>

WILEY-VCH

www.chemphyschem.org

A Journal of



$\text{TCutPNO} = 10^{-7}$, and $\text{TCutMKN} = 10^{-3}$) were used, as recommended for applications targeting the most accurate values.^[61, 75, 77] The following triple- ζ basis sets with diffuse functions were applied to describe hydrogen atom as well as all elements (C, N; O, F) of the second period of the periodic table.^[78] All used atoms of the third period (P, S, Cl) have been described with the cc-pVTZ basis sets of Dunning and Woon.^[79] The correlation fitting basis sets necessary for the resolution of identity approximation as part of the DLPNO scheme, were generated by AutoAux routine developed by Neese and co-workers^[80] as a part of ORCA 4 code.

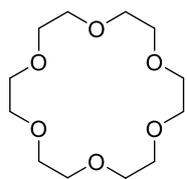
It has to be mentioned that despite DLPNO-CCSD(T)/cc-pVTZ relative conformational energies have been used as references, these still might be contaminated to some extent by intramolecular basis set superposition error (BSSE), especially for the species covered in this study as all DLPNO-CCSD(T)/cc-pVTZ relative conformational energies turned out to be practically identical to their PBE-D3 counterparts, vide infra. In general, the DFT energies are known to be less vulnerable to BSSE, and the dispersion corrected DFT methods have even been recommended for accurate relative conformational energies.^[83]

All XYZ geometries are available in Electronic Supporting Information (ESI) as well as in **37Conf8** collection^[84] on Figshare website. We expect this dataset can be of use for projects relevant to conformational search and/or method development.

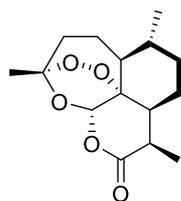
Computationally-Efficient Tested Methods for Conformational Energies

First, classical semiempirical methods MNDO, AM1, PM3 along with their contemporary counterparts RM1,^[37] PM6^[29] and PM7^[32] all as implemented in MOPAC 2016^[85] computer program have been selected due to their wide applications in organic chemistry. For better accuracy in energy evaluation all MOPAC calculations have been accomplished with the keyword `%NOO=0` as we were running with the `\^`, `[!AOT CYME Ad A] !c^` `Ab { } • A A^ [{ d^ A` Second, NDDO-like semiempirical method QM developed by Laikov^[86, 87] to reproduce CCSD(T) data and implemented in Priroda electronic structure package have been included in this work as well. For more reliable SP energies and geometry optimization the following Priroda keywords have been specified: `%M { ' } H A d ! a & M F ^ 6 + A V @ a A & ~] A [- A a) • a A` functional theory tight binding schemes have been also considered: DFTB method of Gaus and co-workers^[33, 88-90] as implemented in DFTB+ software^[91] (DFTB3 method, `SCCTolerance = 5e-7`, `Driver = LBFGS`, `MaxForceComponent = 5e-6`) and Geometry, Frequency, Noncovalent, eXtended Tight Binding (GFN-xTB)^[34] and improved version (GFN2-xTB)^[92] as implemented in `xtb & A C a % [A c a + [a] D a a ! A` the purpose of comparison we included drug design industry appreciated MMFF94 force field method as implemented in TINKER code^[43] `C a % ^, d } + A [- a A and % T U A O : a a) A e e f + A`

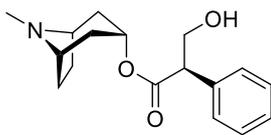
The methods have been tested in reproducing relative conformational energies via SP energy evaluations on geometries obtained in Section 2.2.3 and geometry optimizations thereof.



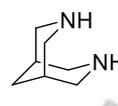
18-crown-6



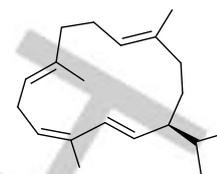
Artemisinin



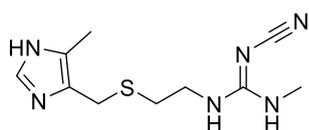
Atropine



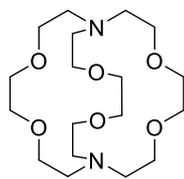
Bispidine



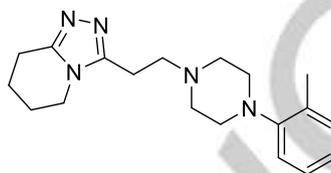
Cembrene



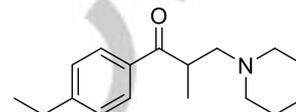
Cimetidine



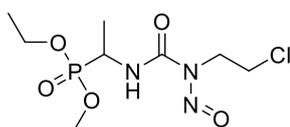
Cryptand-222



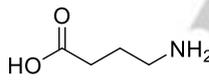
Dapiprazole



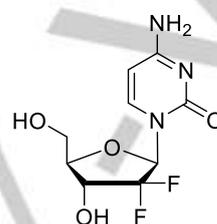
Eperisone



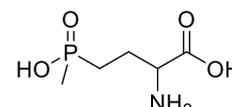
Fotemustin



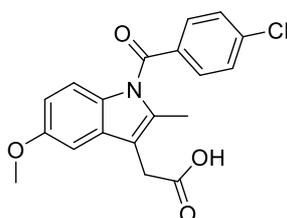
GABA



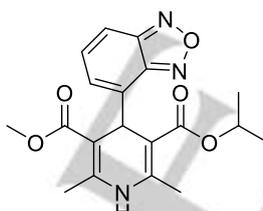
Gemcitabine



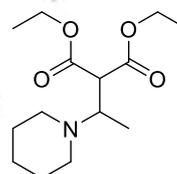
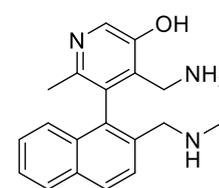
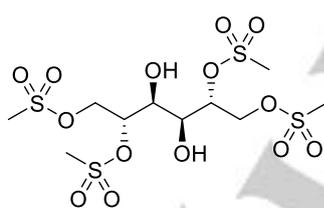
Glufosinate



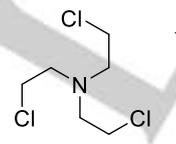
Indometacin



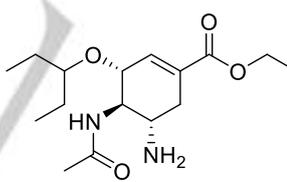
Isradipine

Knoevenagel_cond ^aLiu_catal ^b

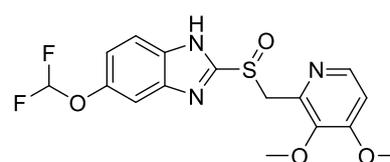
Mannosulfan



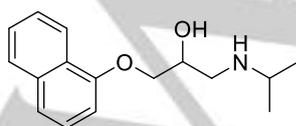
Nitrogen mustard



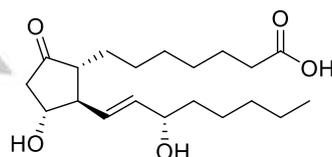
Oseltamivir



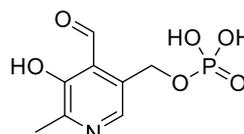
Pantoprazole



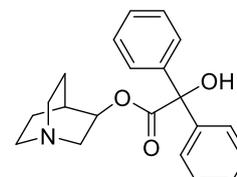
Propranolol



ProstaglandinE1



Pyridoxal phosphate



Quinclidinyl benzilate

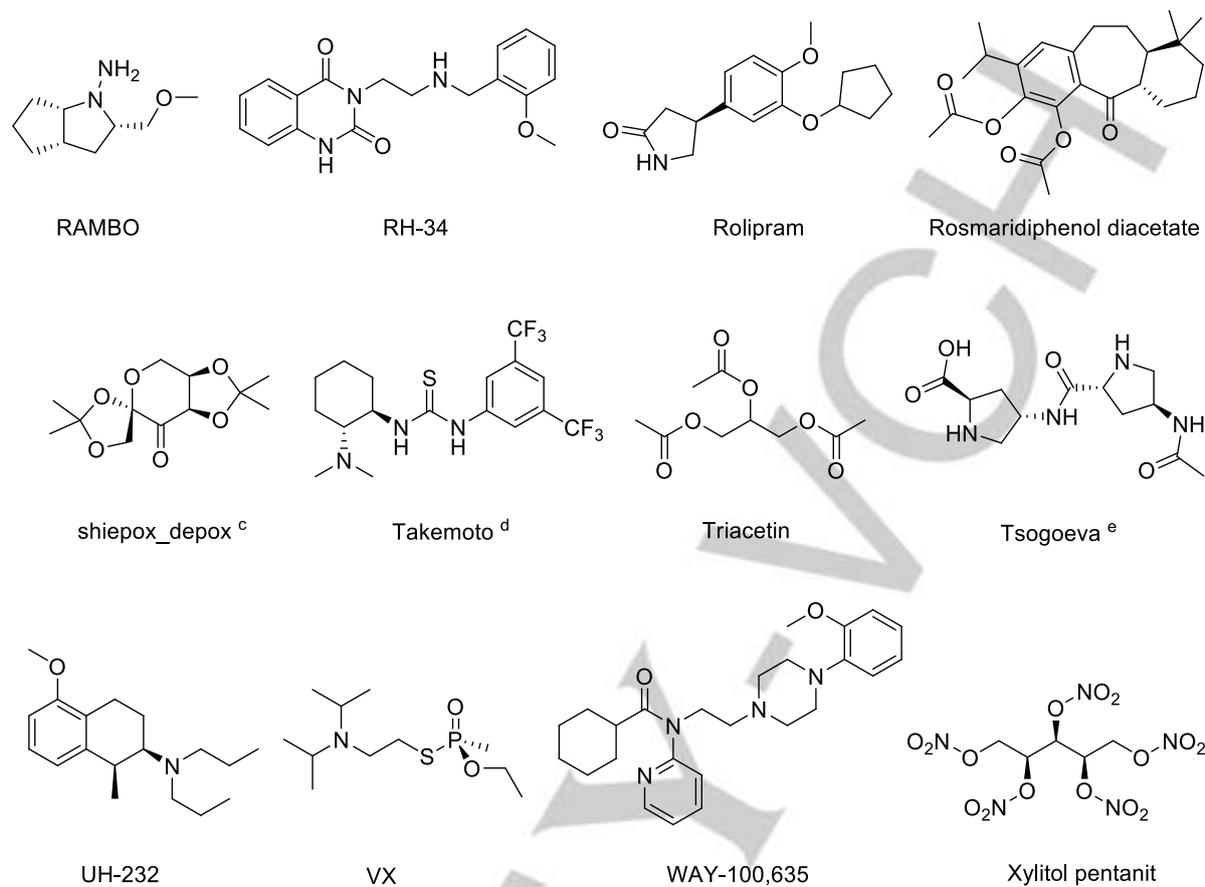


Figure 1. 37Conf8 dataset. a) Intermediate in catalytic cycle of Knoevenagel Condensation, where piperidine plays role of organocatalyst; b) axially chiral [1,10]metallophthalazine derivative; c) catalyst used in Shi epoxidation^[94] d) bifunctional thiourea organocatalyst suggested by Takemoto^[95-97] e) amino-proline based dipeptide used by Tsogoeva as enantioselective organocatalyst^[98]

Comparing the Conformational Energies and Geometries

Calculation of Pearson Correlation Coefficient To quantify the correlation between relative conformational energies from semiempirical computational chemistry methods and MMFF94 with the reference DLPNO-CCSD(T)/cc-pVTZ//PBE-D3/def2-tzvp conformational energies, we have calculated the Pearson correlation coefficient r using the following formula:

$$r = \frac{\sum_{i=1}^n (E_i^X - \bar{E}^X)(E_i^Y - \bar{E}^Y)}{\sqrt{\sum_{i=1}^n (E_i^X - \bar{E}^X)^2 \sum_{i=1}^n (E_i^Y - \bar{E}^Y)^2}} \quad (1)$$

where X is the tested electronic structure theory method to obtain the conformational energies and Y is the method to obtain the reference conformational energies, i.e. DLPNO-CCSD(T)/cc-pVTZ//PBE-D3/def2-tzvp, n is the number of conformations calculated for a given compound, E_i^X is the conformational energy of i^{th} conformer, and \bar{E}^X is the average conformational energy for n conformers from a given compound. r is any value in the interval $[-1, +1]$. If the value is close to 1 there is an absolute correlation, and if this value is close to -1 there is an anticorrelation.

While the r^2 is the most popular criterion in the field of cheminformatics to quantitatively measure the correlation (r^2 describing the quality of the conformational energies from a certain method as it is always positive and cannot distinguish between correlation and anti-correlation. In particular, the large (close to 1) r^2 values can correspond to both correlation ($r^2 = 1$) and anti-correlation ($r^2 = 1$).

Calculation of the Mean Absolute Errors in Conformational Energies Another criterion to judge the quality of the relative conformational energies is to calculate the mean absolute deviation (MAE) between the relative energies obtained with particular method and corresponding reference values (DLPNO-CCSD(T)/cc-pVTZ//PBE-D3/def2-tzvp). The following formula was used to calculate the mean absolute error of particular compound:

$$MAE = \frac{1}{n} \sum_{i=1}^n |E_i^X - E_i^Y| \quad (2)$$

Where X is the tested method, Y is the reference method, n is the number of conformers calculated for the certain compound and E_i is the relative energy of the conformer i .

Calculation of the Mean Absolute Errors in Dihedral Angles of Optimized Conformers Deviations in molecular geometries of the reference conformers and their semiempirical/MMFF94 counterparts can be estimated via different ways. In this work we avoided the most straightforward classical Cartesian coordinates-based root-mean-squared-deviations (RMSD) as it is unsuitable for torsion/dihedral angle in the middle of the molecule results in large RMSD for the atoms far from the torsion/dihedral angle. Instead, we decided to calculate deviation in molecular geometry based on internal parameters, namely torsion angles, see Equation 3. As in the case of labile structures dihedrals are very sensitive to geometry optimization method, we postulated that the structures are most likely correspond to the same conformer if the change in dihedral angle is below 20 degrees. When the difference in corresponding dihedral angle exceeds 60 degrees it

is most likely that we deal with different conformers. Molecular geometries and relative conformational energies resulting from geometry optimizations with semiempirical/FF methods indicate how suitable the particular method for conformational search. The conformational energies derived from the single point energies are aimed at identification of the origin of mistake: either geometry or unrealistic energy of correct (nearly correct) structure.

(3)

Where X is the tested method, Y is the reference method, n is the number of conformers calculated for the certain compound, m is the number of rotatable bonds, D_{ij} is the value of j -th dihedral angle in i -th conformer.

2. Results and discussions

The Results and Discussion is organized as follows. Section 2.1 discusses an ability of the tested methods in reproducing the right order of the SP-based conformational energies on DFT optimized molecular geometries via calculation of the Pearson correlation coefficient. Section 2.2 comments on the performance of tested methods in reproducing of absolute conformational energies calculated on fixed geometries. Conformational changes taking place during geometry optimization with semiempirical and MMFF94 methods are examined in Section 2.3. Practical recommendations for the choice of methods for conformational search and sampling in flexible organic molecules are given.

2.1 Correctness of an Order of Conformational Energies (Pearson Correlation Coefficient)

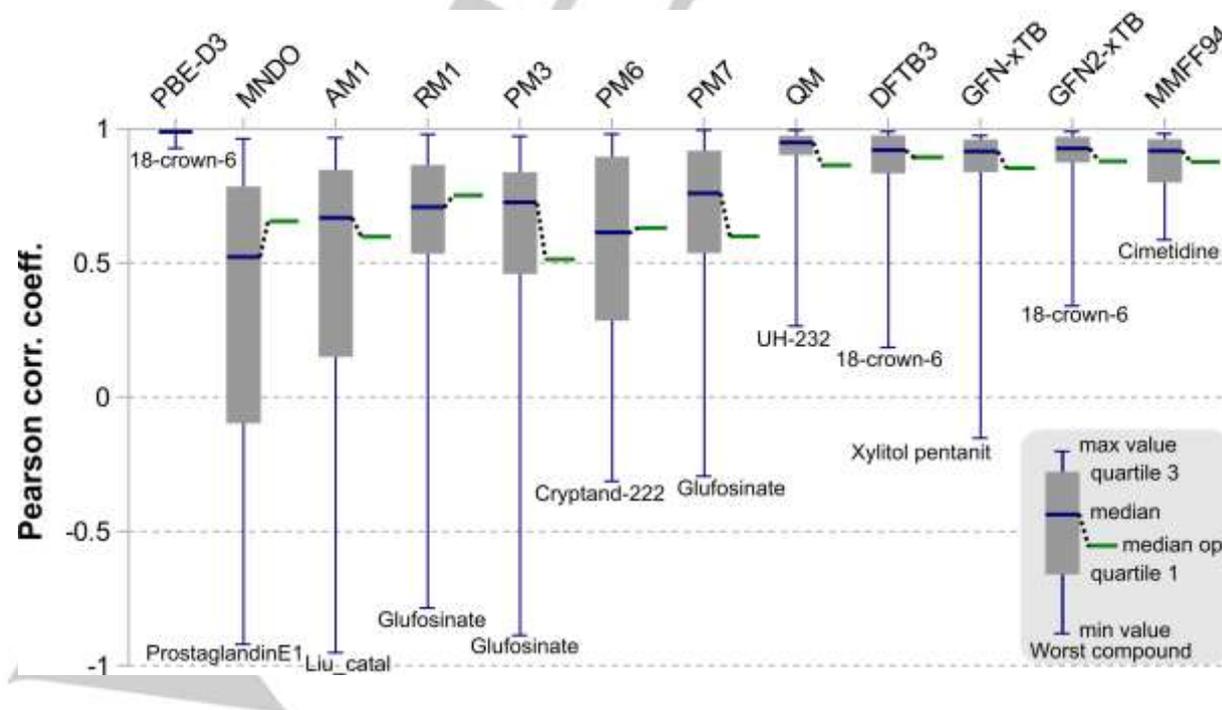
As the semiempirical and force field methods are mostly used for pre-selection of the relevant low-energies conformers to be studied later with more accurate DFT and WFT methods, we first tested the quality of relative conformational energies. Pearson correlation coefficient r is a good indicator of the ability to reproduce the correct order of conformational energies, see Figure 2. An examination of Figure 2 reveals DFT PBE-D3 method with median of $r = 0.99$ to be the best performer clearly standing out from other approximations tested in this work. Even the worst correlation coefficient obtained for 18-crown-6 compound turned out to be 0.93 which is acceptable. These results merged with our previous findings indicate DFT to be a reliable tool to study conformational space of organic and inorganic molecules. Other tested methods clearly form the two separate groups based on their median values of r and quartile 1 (Q1) - quartile 3 (Q3) span.

The first group of methods demonstrating rather moderate performance are the semiempirical and force field methods. The methods within this group appear in the order of their historical development. The MNDO scheme (1977) shows the worst performance with $r = 0.52$ and Q1/Q3 values of $-0.10/0.79$. Somewhat better results, namely r of 0.67 and Q1/Q3 of $0.15/0.85$, were obtained for AM1 (1985) method. Finally, even better r of 0.71 and Q1/Q3 values of $0.53/0.87$ were documented for RM1 method which is a further improvement of AM1, illustrating a clear progress in method development. Another successful reparametrization of MNDO scheme, PM3 method of Stewart (1989) resulted in r of 0.73 and Q1/Q3 of $0.46/0.84$ which

FULL PAPER

is slightly better comparing to its competitor AM1. PM6 (2007) method with $r = 0.61$ and Q1/Q3 values of 0.29/0.90 demonstrated slightly worse performance comparing to its predecessor PM3. Perhaps, it is compensated by the fact that PM3 is applicable only to molecules comprised only of a few popular in organic chemistry elements while PM6 can be used for most elements of the Periodic Table. Finally, for PM7 (2013) method, a successor of PM6, yet quite moderate $r = 0.76$ and Q1/Q3 = 0.54/0.92 values were obtained. Coupled with the fact that practically for every method from this group large negative values ranging from -0.95 (MNDO) to -0.29 (PM7) have been obtained implying σ - π interactions in σ - π based schemes should be utilized with great care and caution for conformational search/sampling-related tasks. It has to be stressed that even 0.70 correlation is quite poor and results of such quality should be mainly used for trend identification or qualitative estimations. At the same time correlation coefficient of 0.1-0.3 observed for a few combinations of methods/compounds indicates that predicted conformations are not reliable. In addition to SP-based conformational energies we also obtained their full geometry optimization-based counterparts. The conformational energies derived from consistently optimized geometries did not demonstrate any better Pearson correlation coefficient values, see Figure 2.

The second performance-based group of methods can be relatively computationally expensive approximations QM,^[86, 87] DFTB,^[33, 88-90] GFN-xTB,^[34] and GFN2-xTB^[92]. All these approximations resulted in relatively high median values for **37conf8** dataset in range 0.92 (GFN-xTB) . 0.95 (QM) and acceptable Q1 - Q3 values in range 0.83 (DFTB, Q1) . 0.98 (DFTB, Q3). Computationally inexpensive force field method MMFF94^[39-41] also belongs to this group over-performing in contradiction with the results of ref.^[24] We find an origin of this discrepancy in principal difference in selection of conformers for the study. While Hutchison et al. focused exclusively on low-lying conformers (according to PM7 energies) we worked with the conformers forming relatively wide energy interval. Our tests carried on conformers forming small energy interval reaffirm low correlation between MMFF94 and DFT. However, the correlation recovers when higher lying conformers are taken into account. Despite MMFF94 provides comparable to other methods from this group performance, it has to be noted that it is applicable to significantly more narrow classes of molecules, and its extension is definitely not a straightforward task. In addition to SP-based conformational energies we also obtained their counterparts based on fully reoptimized geometries with the same methods. Thus derived conformational energies did not show any significant improvements over their SP-based counterparts, see Figure 2. As can be seen from Figure 2, consistent optimization with all the methods from the second group resulted in similar median values of a Pearson correlation coefficient. In general, based on the obtained results we recommend the methods from the second group to be used for reliable ranking of the conformational energies. However, caution is still needed as even for these methods unexpected failures have been documented, see below.



FULL PAPER

Figure 2. Pearson correlation coefficients (medians, quartile 1 (Q1), quartile 3 (Q3), highest and smallest values) obtained for correlation between Single-Point (SP) energy based DFT, semiempirical and force field conformational energies and their reference DLPNO-CCSD(T)/cc-pVTZ counterparts. Median values after consistent geometry optimization with all but PBE-D3 methods are also provided.

2.2 Correctness of Absolute Conformational Energies

An ability of all tested methods in reproducing of absolute conformational energies is shown in Figure 3. As expected, the DFT PBE-D3 method provided the most accurate conformational energies with median conformational energy error of only 0.6 kcal/mol. Similarly to what was observed earlier for Pearson correlation coefficients, the methods clearly form the two separate groups. Again, the first group of semiempirical methods. Thus, the largest median error of 3.5 kcal/mol was obtained for MNDO method. Its successors, AM1 and PM3 methods have resulted in smaller error median values of 2.9 and 3.0 kcal/mol, correspondingly. Q1-Q3 interval comparing to MNDO method has been significantly reduced and amounted to 1.9-4.2 kcal/mol for AM1 and 1.8-3.9 kcal/mol for PM3. Even smaller median error of 2.5 kcal/mol and Q1/Q3 values of 1.6 and 3.1 kcal/mol have been obtained for RM1 method. PM6 approach resulted in slightly larger median of 2.7

kcal/mol which comes at price of broader applicability comparing to RM1 parameterized for a limited set of elements. Finally, PM7 method provided median error of 2.3 kcal/mol.

The second group of methods with clearly better performance is represented by contemporary semiempirical methods. The QM method of Laikov resulted in lowest median error of 1.4 kcal/mol, followed closely by GFN2-xTB with error of 1.5 kcal/mol. Slightly larger median errors of 1.7 and 1.8 kcal/mol were obtained for DFTB3 and GFN-xTB methods, correspondingly. In contrast to values were obtained for contemporary semiempirical methods from the second group.

The most striking difference comparing to the Pearson correlation coefficient-based results is grouping of the force field MMFF94 method. Its median error of 2.7 kcal/mol practically matches that of PM6 method of Stewart. It has also to be noted that conformational energies derived from reoptimized with the same semiempirical or MMFF94 method geometries did not lead to any statistically sound changes, see Figure 3.

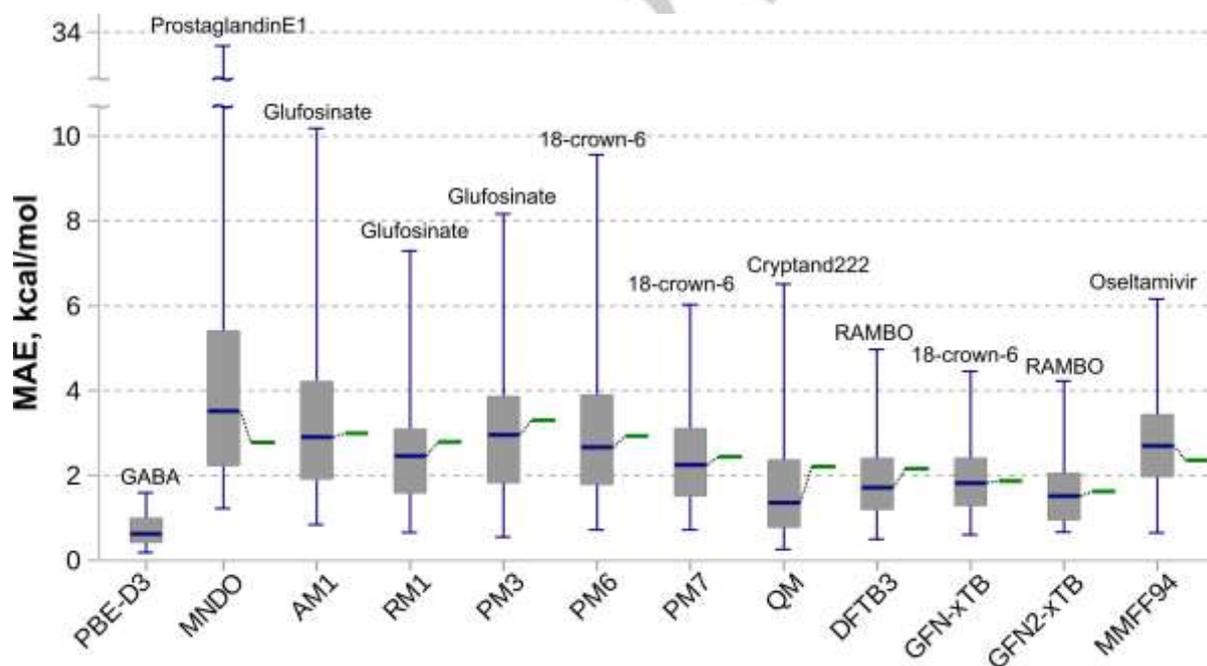


Figure 3. Absolute errors (medians, quartile 1 (Q1), quartile 3 (Q3) and highest/smallest mean absolute error values) obtained for Single-Point (SP) energy based DFT, semiempirical and force field conformational energies with respect to their reference DLPNO-CCSD(T)/cc-pVTZ counterparts. Median values after consistent optimization with all but PBE-D3 methods are also provided.

Finally, it has to be stressed that despite good to excellent performance found in general for contemporary semiempirical methods, some spectacular failures are still possible. Thus, based on Figure 2 18-crown-6 turned out to be the most problematic molecule for DFT, DFTB, GFN2-xTB and the second most problematic for GFN-xTB methods (see SI). As this compound is comprised of only C, H, O atoms, bad performance of contemporary semiempirical methods is rather puzzling. Since 18-crown-6 has big number of conformers (>3136 according to ref.^[99]

or ca. 31385 according to our data Tinker scan MMFF94 estimation with an energy window of 20 kcal/mol). For 37Conf8 dataset we have selected only 8 most stable conformers representing following symmetry point groups: S_6 , C_1 , D_{3d} , C_2 , C_s , C_3 , D_2 , C_{2h} . An analysis indicated that the largest difference between conformational energies obtained with all tested methods and DLPNO-CCSD(T) reference values have been detected for D_{3d} - and D_2 -symmetry conformers (see SI). To illustrate these failures, in Figure 4 the energy

FULL PAPER

difference between energies of D_{3d} , conformers and and D_2/S_6 conformers is depicted.

WILEY-VCH

Accepted Manuscript

FULL PAPER

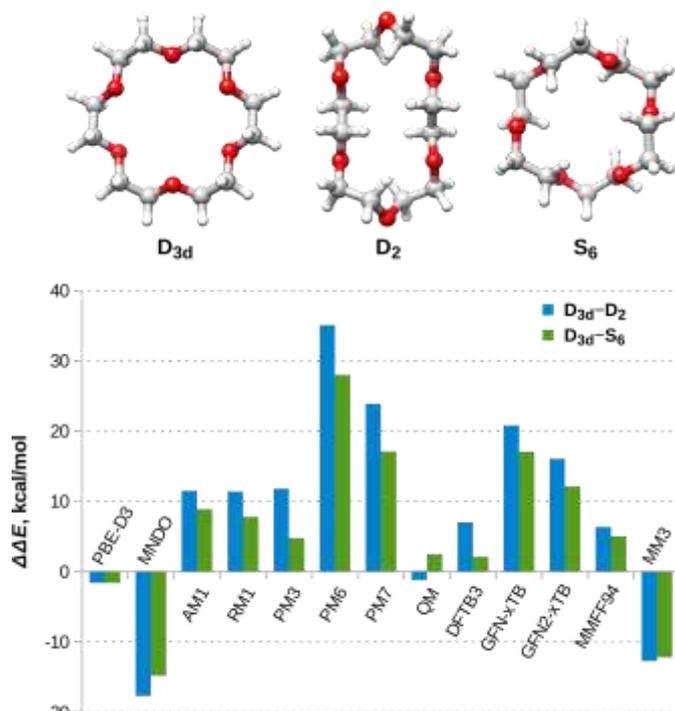


Figure 4. Energy difference

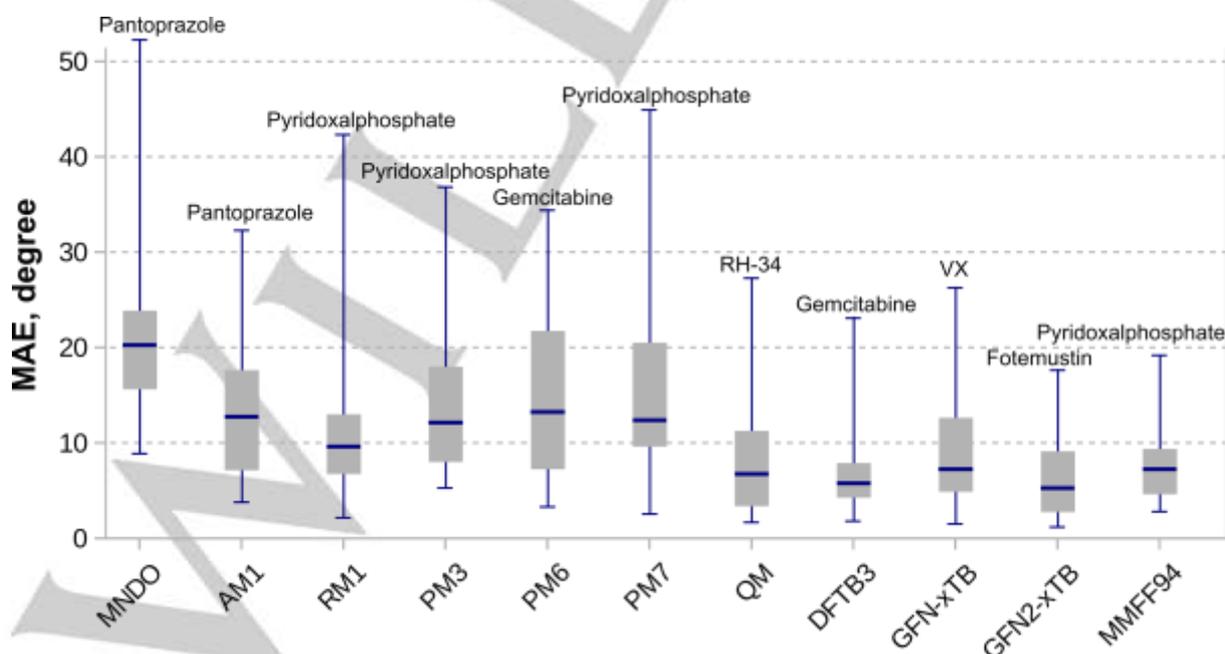
between D_{3d} and D_2/S_6 18-crown-6 conformers obtained for all methods tested based on SP energies on PBE-D3 optimized geometries.

The largest errors in the conformational energies difference have been obtained for PM6 method and amounted to 35.1 for energy

difference between D_{3d} and D_2 and 28.0 kcal/mol between D_{3d} and S_6 conformer. PM7 and GFN-xTB methods turned out not to be significantly better and both resulted in errors about 20 kcal/mol. For modified method of Grimme and co-workers, GFN2-xTB^[92], smaller but yet non-negligible deviations of 12-16 kcal/mol have been detected indicating certain success in method development. The PBE-D3 and QM methods provided the smallest errors and are recommended for accurate energies of 18-crown-6 conformational energies. It has to be noted that difficulties in reproducing conformational energies of 18-crown-6 have been detected earlier. Thus, in ref.^[99] MM3 force field failed in reproducing D_{3d} . S_6 reference conformational energies difference obtained with MP2/6-31+G* approximation. Wrong description of hydrogen bonds in MM3 was suggested to be an origin. It is encouraging, however, that MMFF94 resulted in much smaller errors. We suggest these differences in energies of 18-crown-6 conformers to be utilized in training and testing of semiempirical and force field method.

2.3 Conformational Changes upon Geometry Optimization

To see whether the inconsistencies observed in Sections 1 and 2 are originated from the fundamental differences in conformational potential energy surfaces, we optimized all the structures self-consistently with tested semiempirical and force field methods. The conformational changes upon geometry optimization, i.e. differences in dihedral angles corresponding to rotatable bonds with respect to their reference DFT values are given in Figure 5.



FULL PAPER

Figure 5. Calculated errors in the values of dihedral angles (degrees) corresponding to rotatable bonds obtained upon geometry optimizations with all but PBE-D3 tested methods.

WILEY-VCH

Accepted Manuscript

FULL PAPER

Similarly to what was found in Section 3.1 the methods for the two groups based on their performance in reproducing PBE-D3 resulted in relatively high median values of an average error in dihedral values ranging from 20° (MNDO) to 10° (RM1). The second group of methods providing substantially smaller errors for dihedral angles includes contemporary semiempirical methods. The medians of errors in dihedrals lasts from 7° (QM, GFN-xTB) to 5° (GFN2-xTB). Force field MMFF94 method resulted in similar to the second group of methods performance with the median of an average dihedral angle error of 7°.

Taken together, the results obtained in this work suggest the following protocol for conformational search and sampling in flexible organic molecules. An optimization of molecular structures as well as pre-selection of the relevant low-lying conformers can be done with computationally cheap MMFF94 approximation. Afterwards, the absolute conformational energies of the energetically low-lying structures need for accurate sampling to average the properties over the conformational space is to be done with one of the contemporary semiempirical methods. An obvious limitation of this strategy is its non-applicability to systems that contain transition metals and other poorly parametrized elements. Moreover, it has to be stressed that even for purely organic systems no strategy based on semiempirical/molecular mechanics methods is guaranteed to be free of unexpected failures as it was observed for conformers of 18-crown-6 in this work. On the other hand, DFT methods are expected to be more robust for the conformational search however these become prohibitively expensive upon growth of the number of rotatable bonds.

Conclusions

In this work we developed the **37conf8** database, which contains eight accurate DLPNO-CCSD(T)/cc-pVTZ//PBE-D3/def2-tzvp conformational energies and spatial structures of 37 organic molecules representing molecules of industrial interest and having a large variety of different topological and functional group features. The **37conf8** database has been developed using DLPNO-CCSD(T)/cc-pVTZ//PBE-D3/def2-tzvp, GFN2-xTB, GFN2-xTB methods as well as the MMFF94 force field approximation. Contemporary semiempirical methods have been found to outperform their standard counterparts sorting conformers in the right order and providing reliable absolute conformational energies and spatial structures in the vast majority of cases, although at significantly higher computational costs. MMFF94 force field approximation groups together with contemporary semiempirical methods in reproducing the correct order of conformational energies and spatial structures of the conformers. However, an error in absolute energies derived from MMFF94 protocol is relatively large and closer to that obtained for standard semiempirical methods. Hence, for routine conformational search and sampling in flexible organic molecules of realistic size we recommend the following two-step strategy: a) geometry optimization and pre-selection of the relevant (energetically low lying) conformers with the computationally inexpensive MMFF94 force field approximation to sample exhaustively the conformational space in reasonable time; b) single point energies needed to derive the *absolute*

conformational energy to build realistic Boltzmann-weighted ensemble of conformations with contemporary semiempirical methods. However, even this strategy does not guarantee reliable conformational search and sampling due to unexpected failures of contemporary semiempirical methods revealed in this work for conformers of some compounds (e.g. D_{3d} and D_2 18-crown-6 conformers) and/or absence/poor parametrization of some elements as transition metals in tested approximations. The density functional theory methods have been found to outperform all other tested schemes and can be recommended for reliable conformational search in organic systems with ca. 100 atoms and 500-1000 unique conformers as suggested ref.^[14] In reality, a dominant part of important organic/organometallic molecules will possess much larger number of conformers. Thus, a total number of unique conformers have been estimated to be 4217 for arginine dipeptide^[100] with all found conformers to be within 33 kcal/mol and the energy threshold for conformational sampling of 92 kcal/mol. For 18-crown-6 we have found 31385 conformers in this work by procedure describe in the Methodology and Computational Details Section and an energy window of 20 kcal/mol. By systematic rotation of each rotatable bond by 90° degrees and rejecting the conformers with serious steric repulsion more than 98277, 294624, 1327104, 2301664 and 21233664 conformers have been obtained for tetrapeptides GGGG, GVGG, GTGG, GGYG and GSDG, correspondingly.^[101] It makes all existing DFT approximations prohibitively expensive for an exhaustive search/sampling. Hence, we believe finding/developing of computationally fast computational protocol providing reliable conformational energies forming a basis of conformational search and sampling is of paramount importance and derived **37conf8** database is likely to facilitate this process.

Conflicts of interest

There are no conflicts to declare

Acknowledgements

The research reported in this publication was supported by funding from the Government of the Russian Federation (grant number 14-02-2018-286). L.C. gratefully acknowledges the financial support from King Abdullah University of Science and Technology (KAUST). A. G. gratefully acknowledges the financial support from KAUST (grant number 14-03-00564). For computer time, this research used the resources of the Supercomputing Laboratory at King Abdullah University of Science and Technology (KAUST) in Thuwal, Saudi Arabia and Novosibirsk State University Supercomputer Center (NUSC) in Novosibirsk, the Russian Federation.

Keywords

Conformational energies, density functional theory, DLPNO-CCSD(T), force field, semiempirical methods

Supporting Information

Electronic Supplementary Information (ESI) available: Cartesian coordinates (Å) of PBE-D3/def2-tvp optimized structures and DLPNO-CCSD(T)/cc-pVTZ, PBE-D3/def2-tvp, MNDO, AM1, RM1, PM3, PM6, PM7, QM, DFTB, GFN-xTB, GFN2-xTB, MMFF94 single-point energies on these structures, MNDO, AM1, RM1, PM3, PM6, PM7, QM, DFTB, GFN-xTB, GFN2-xTB,

FULL PAPER

MMFF94 re-optimized molecular geometries as well as corresponding energies, tabulated values forming the basis of Figures 2-5, tabulated conformational energies and values of dihedral angles corresponding to the rotatable bonds. This material is available free of charge.

Notes and references

- [1] K. T. Butler, F. J. Luque, X. Barril *J. Comput. Chem.* **2009**, *30*, 601-610.
- [2] T. Hayami, K. Kasahara, H. Nakamura, J. Higo *Journal of Computational Chemistry*. **2018**, *0*.
- [3] C. Parish, R. Lombardi, K. Sinclair, E. Smith, A. Goldberg, M. Rappleye, M. Dure *J. Mol. Graphics Modell.* **2002**, *21*, 129-150.
- [4] P. Labute *J. Chem. Inf. Model.* **2010**, *50*, 792-800.
- [5] A. G. Csaszar *J. Am. Chem. Soc.* **1992**, *114*, 9568-9575.
- [6] A. L. East, Z. F. Liu, C. McCague, K. Cheng, J. S. Tse *J. Phys. Chem. A*. **1998**, *102*, 10903-10911.
- [7] A. Salam, M. S. Deleuze *J. Chem. Phys.* **2002**, *116*, 1296-1302.
- [8] S. Tsuzuki, A. A. Arai, K. Nishikawa *J. Phys. Chem. B*. **2008**, *112*, 7739-7747.
- [9] F. Ramirez, C. Z. Hadad, D. Guerra, J. David, A. Restrepo *Chem. Phys. Lett.* **2011**, *507*, 229-233.
- [10] S. Kozuch, S. M. Bachrach, J. M. L. Martin *J. Phys. Chem. A*. **2014**, *118*, 293-303.
- [11] N. V. Tukachev, V. A. Bataev, I. A. Godunov *J. Quant. Spectrosc. Radiat. Transfer*. **2017**, *192*, 30-41.
- [12] K. N. Kirschner, W. Heiden, D. Reith *Mol. Phys.* **2017**, *115*, 1155-1165.
- [13] M. Kumar, J. S. Francisco *J. Phys. Chem. A*. **2017**, *121*, 6652-6659.
- [14] J. Rezac, D. Bim, O. Gutten, L. Rulisek *J. Chem. Theory Comput.* **2018**, *14*, 1254-1266.
- [15] N. D. Kondratyuk, G. E. Norman, V. V. Stegailov *Polym. Sci., Ser. A*. **2016**, *58*, 825-836.
- [16] N. D. Kondratyuk, G. E. Norman, V. V. Stegailov *J. Chem. Phys.* **2016**, *145*, 204504/204501-204504/204509.
- [17] A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard, III, W. M. Skiff *J. Am. Chem. Soc.* **1992**, *114*, 10024-10035.
- [18] C. J. Casewit, K. S. Colwell, A. K. Rappe *J. Am. Chem. Soc.* **1992**, *114*, 10035-10046.
- [19] K. Gundertofte, T. Liljefors, P.-o. Norrby, I. Pettersson *J. Comput. Chem.* **1996**, *17*, 429-449.
- [20] I. Pettersson, T. Liljefors *Rev. Comput. Chem.* **1996**, *9*, 167-189.
- [21] V. V. Stegailov, N. D. Orekhov, G. S. Smirnov in HPC Hardware Efficiency for Quantum and Classical Molecular Dynamics, Vol. (Ed. A. Eds.: Editor), Springer International Publishing, City, **2015**, pp.469-473.
- [22] A. Mazzanti, D. Casarini *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2*, 613-641.
- [23] A. Metsala, S. Tamp, K. Danilas, U. Lille, L. Villo, S. Vija, T. Pehk, O. Parve *J. Theor. Chem.* **2014**, *714164/714161-714164/714110*, 714111 pp.
- [24] I. Y. Kanal, J. A. Keith, G. R. Hutchison *Int. J. Quantum Chem.* **2018**, *118*, n/a.
- [25] A. T. Cavin, A. Hillisch, F. Uellendahl, S. Schneckener, A. H. Goeller *J. Chem. Inf. Model.* **2018**, *58*, 1005-1020.
- [26] D. Gruzman, A. Karton, J. M. L. Martin *J. Phys. Chem. A*. **2009**, *113*, 11974-11983.
- [27] U. R. Fogueri, S. Kozuch, A. Karton, J. M. L. Martin *J. Phys. Chem. A*. **2013**, *117*, 2269-2277.
- [28] R. Bjornsson, I. Arnason *Phys. Chem. Chem. Phys.* **2009**, *11*, 8689-8697.
- [29] J. J. P. Stewart *J. Mol. Model.* **2007**, *13*, 1173-1213.
- [30] J. Rezac, J. Fanfrlik, D. Salahub, P. Hobza *J. Chem. Theory Comput.* **2009**, *5*, 1749-1760.
- [31] J. Rezac, P. Hobza *J. Chem. Theory Comput.* **2012**, *8*, 141-151.
- [32] J. J. P. Stewart *J. Mol. Model.* **2013**, *19*, 1-32.
- [33] M. Gaus, Q. Cui, M. Elstner *J. Chem. Theory Comput.* **2011**, *7*, 931-948.
- [34] S. Grimme, C. Bannwarth, P. Shushkov *J. Chem. Theory Comput.* **2017**, *13*, 1989-2009.
- [35] W. Weber, W. Thiel *Theor. Chem. Acc.* **2000**, *103*, 495-506.
- [36] T. Tuttle, W. Thiel *Phys. Chem. Chem. Phys.* **2008**, *10*, 2159-2166.
- [37] G. B. Rocha, R. O. Freire, A. M. Simas, J. J. P. Stewart *J. Comput. Chem.* **2006**, *27*, 1101-1111.
- [38] B. R. White, C. R. Wagner, D. G. Truhlar, E. A. Amin *J. Chem. Theory Comput.* **2008**, *4*, 1718-1732.
- [39] T. A. Halgren *J. Comput. Chem.* **1996**, *17*, 490-519.
- [40] T. A. Halgren *J. Comput. Chem.* **1996**, *17*, 520-552.
- [41] T. A. Halgren *J. Comput. Chem.* **1996**, *17*, 553-586.
- [42] Y. Minenkov, D. I. Sharapa, L. Cavallo *J. Chem. Theory Comput.* **2018**, *14*, 3428-3439.
- [43] J. A. Rackers, Z. Wang, C. Lu, M. L. Laury, L. Lagardère, M. J. Schnieders, J.-P. Piquemal, P. Ren, J. W. Ponder *Journal of Chemical Theory and Computation*. **2018**, *14*, 5273-5289.
- [44] N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, G. R. Hutchison *J. Cheminf.* **2011**, *3*, 33.
- [45] P. Tosco, T. Balle, F. Shirir *J. Mol. Model.* **2011**, *17*, 3021-3023.
- [46] A. M. Gillespie, G. R. Morello, D. P. White *Organometallics*. **2002**, *21*, 3913-3921.
- [47] J. P. Perdew, K. Burke, M. Ernzerhof *Phys. Rev. Lett.* **1996**, *77*, 3865-3868.
- [48] J. P. Perdew, K. Burke, M. Ernzerhof *Phys. Rev. Lett.* **1997**, *78*, 1396.
- [49] D. N. Laikov *Chem. Phys. Lett.* **2005**, *416*, 116-120.
- [50] D. N. Laikov, Y. A. Ustynyuk *Russ. Chem. Bull.* **2005**, *54*, 820-826.
- [51] J. Rezac, K. E. Riley, P. Hobza *J. Chem. Theory Comput.* **2011**, *7*, 2427-2438.
- [52] J. Rezac, K. E. Riley, P. Hobza *J. Chem. Theory Comput.* **2014**, *10*, 1359-1360.
- [53] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, Williams, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, D. J. Fox in Gaussian 16 Rev. A.01, Vol. (Ed. A. Eds.: Editor), City, **2016**.
- [54] Y. Minenkov, L. Cavallo *ACS Omega*. **2017**, *2*, 8373-8387.
- [55] S. Grimme, J. Antony, S. Ehrlich, H. Krieg *J. Chem. Phys.* **2010**, *132*, 154104/154101-154104/154119.
- [56] S. Grimme, S. Ehrlich, L. Goerigk *J. Comput. Chem.* **2011**, *32*, 1456-1465.
- [57] Y. Minenkov, A. Singstad, G. Occhipinti, V. R. Jensen *Dalton Trans.* **2012**, *41*, 5526-5541.
- [58] F. Weigend, R. Ahlrichs *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297-3305.
- [59] C. Riplinger, F. Neese *J. Chem. Phys.* **2013**, *138*, 034106/034101-034106/034118.
- [60] C. Riplinger, B. Sandhoefer, A. Hansen, F. Neese *J. Chem. Phys.* **2013**, *139*, 134101/134101-134101/134113.
- [61] D. G. Liakos, M. Sparta, M. K. Kesharwani, J. M. L. Martin, F. Neese *J. Chem. Theory Comput.* **2015**, *11*, 1525-1539.
- [62] D. G. Liakos, F. Neese *J. Chem. Theory Comput.* **2015**, *11*, 4054-4063.
- [63] P. Pinski, C. Riplinger, E. F. Valeev, F. Neese *J. Chem. Phys.* **2015**, *143*, 034108/034101-034108/034117.
- [64] Y. Guo, K. Sivalingham, E. F. Valeev, F. Neese *J. Chem. Phys.* **2016**, *144*, 094111/094111-094111/094116.
- [65] F. Pavosevic, P. Pinski, C. Riplinger, F. Neese, E. F. Valeev *J. Chem. Phys.* **2016**, *144*, 144109/144101-144109/144113.
- [66] C. Riplinger, P. Pinski, U. Becker, E. F. Valeev, F. Neese *J. Chem. Phys.* **2016**, *144*, 024109/024101-024109/024110.
- [67] F. Pavosevic, C. Peng, P. Pinski, C. Riplinger, F. Neese, E. F. Valeev *J. Chem. Phys.* **2017**, *146*, 174108/174101-174108/174110.
- [68] M. Saitow, U. Becker, C. Riplinger, E. F. Valeev, F. Neese *J. Chem. Phys.* **2017**, *146*, 164105/164101-164105/164131.

FULL PAPER

- [69] F. Neese *Wiley Interdisciplinary Reviews: Computational Molecular Science*. **2017**, *8*, e1327.
- [70] F. Neese *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2*, 73-78.
- [71] Y. Minenkov, E. Chermak, L. Cavallo *J. Chem. Theory Comput.* **2015**, *11*, 4664-4676.
- [72] Y. Minenkov, H. Wang, Z. Wang, S. M. Sarathy, L. Cavallo *J. Chem. Theory Comput.* **2017**, *13*, 3537-3560.
- [73] Y. Minenkov, E. Chermak, L. Cavallo *J. Chem. Theory Comput.* **2016**, *12*, 1542-1560.
- [74] Y. Minenkov, V. V. Sliznev, L. Cavallo *Inorg. Chem.* **2017**, *56*, 1386-1401.
- [75] Y. Minenkov, G. Bistoni, C. Riplinger, A. A. Auer, F. Neese, L. Cavallo *Phys. Chem. Chem. Phys.* **2017**, *19*, 9374-9391.
- [76] G. Bistoni, C. Riplinger, Y. Minenkov, L. Cavallo, A. A. Auer, F. Neese *J. Chem. Theory Comput.* **2017**, *13*, 3220-3227.
- [77] Y. Guo, C. Riplinger, U. Becker, D. G. Liakos, Y. Minenkov, L. Cavallo, F. Neese *J. Chem. Phys.* **2018**, *148*, 011101/011101-011101/011105.
- [78] T. H. Dunning, Jr. *J. Chem. Phys.* **1989**, *90*, 1007-1023.
- [79] D. E. Woon, T. H. Dunning, Jr. *J. Chem. Phys.* **1993**, *98*, 1358-1371.
- [80] G. L. Stoychev, A. A. Auer, F. Neese *J. Chem. Theory Comput.* **2017**, *13*, 554-562.
- [81] T. van Mourik, P. G. Karamertzanis, S. L. Price *The Journal of Physical Chemistry A*. **2006**, *110*, 8-12.
- [82] R. Hameed, A. Khan, T. van Mourik *Molecular Physics*. **2018**, *116*, 1236-1244.
- [83] Z. H. Park, S. E. Kim, I. A. Jung, I. A. Jung, A. J. Cohen, A. J. Cohen, A. J. Cohen, A. J. Cohen *Journal of Computational Chemistry*. **2007**, *29*, 861-870.
- [84] D. Sharapa, Y. Minenkov, A. Genaev in 37Conf8. figshare. Collection., Vol. (Ed.^Eds.: Editor), City, **2018**.
- [85] J. J. P. Stewart in MOPAC2016, Vol. (Ed.^Eds.: Editor), Stewart Computational Chemistry, Colorado Springs, CO, USA, City, **2016**.
- [86] D. N. Laikov *J. Chem. Phys.* **2011**, *135*, 134120/134121-134120/134110.
- [87] K. R. Briling *J. Chem. Phys.* **2017**, *147*, 157101/157101.
- [88] M. Gaus, A. Goez, M. Elstner *J. Chem. Theory Comput.* **2013**, *9*, 338-354.
- [89] M. Gaus, X. Lu, M. Elstner, Q. Cui *J. Chem. Theory Comput.* **2014**, *10*, 1518-1537.
- [90] M. Kubillus, T. Kubar, M. Gaus, J. Rezac, M. Elstner *J. Chem. Theory Comput.* **2015**, *11*, 332-342.
- [91] B. Aradi, B. Hourahine, T. Frauenheim *J. Phys. Chem. A*. **2007**, *111*, 5678-5684.
- [92] C. Bannwarth, S. Ehlert, S. Grimme *ChemRxiv. Preprint*. **2018**.
- [93] Y. E. Liu, Z. Lu, B. Li, J. Tian, F. Liu, J. Zhao, C. Hou, Y. Li, L. Niu, B. Zhao *J. Am. Chem. Soc.* **2016**, *138*, 10730-10733.
- [94] Y. Shi, O. A. Wong in 1,2:4,5-Di-O-isopropylidene- -D-erythrohexo-2,3-diulo-2,6-pyranose, Vol. (Ed.^Eds.: Editor), John Wiley & Sons Ltd., City, **2013**, pp.221-224.
- [95] T. Okino, Y. Hoashi, Y. Takemoto *J. Am. Chem. Soc.* **2003**, *125*, 12672-12673.
- [96] T. Okino, S. Nakamura, T. Furukawa, Y. Takemoto *Org. Lett.* **2004**, *6*, 625-627.
- [97] T. Okino, Y. Hoashi, T. Furukawa, X. Xu, Y. Takemoto *J. Am. Chem. Soc.* **2005**, *127*, 119-125.
- [98] S. B. Tsogoeva, S. B. Jagtap, Z. A. Ardemasova *Tetrahedron: Asymmetry*. **2006**, *17*, 989-992.
- [99] N. A. Al-Jallal, A. A. Al-Kahtani, A. A. El-Azhary *The Journal of Physical Chemistry A*. **2005**, *109*, 3694-3703.
- [100] M. Ropo, M. Schneider, C. Baldauf, V. Blum *Sci. Data*. **2016**, *3*, 160009.
- [101] B. Yang, Z. Lin *Comput. Theor. Chem.* **2017**, *1108*, 40-49.

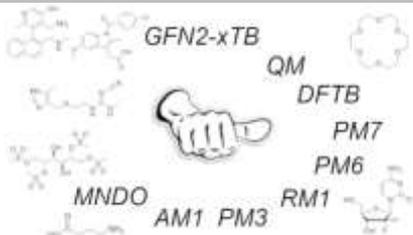
FULL PAPER

Entry for the Table of Contents (Please choose one layout)

Layout 1:

FULL PAPER

Several standard and state-of-the-art semiempirical methods along with the MMFF94 force field approximation have been calibrated towards DLPNO-CCSD(T) conformational energies and spatial structures for 37 organic molecules representing pharmaceuticals, drugs, catalysts, synthetic precursors, industry-related chemicals forming 37conf8 database.



*Dmitry I. Sharapa, Alexander Genaev,
Luigi Cavallo* and Yury Minenkov**

Page No. – Page No.
**A Robust and Cost-Efficient Scheme
for Accurate Conformational
Energies of Organic Molecules**

WILEY-VCH

Accepted Manuscript