FULL PAPER

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

Dmitry I. Sharapa[a], Alexander Genaev[b], Luigi Cavallo*,[c] and Yury Minenkov*,[d]

Abstract: Several standard semiempirical methods as well as the MMFF94 force field approximation have been tested in reproducing 8 DLPNO-CCSD(T)/cc-pVTZ level conformational energies and spatial structures for 37 organic molecules representing pharmaceuticals, drugs, catalysts, synthetic precursors, industry-related chemicals (37conf8 database). All contemporary semiempirical methods surpass their standard counterparts resulting in more reliable conformational energies and spatial structures, even though at significantly higher computational costs. However, even these methods show unexpected failures in reproducing energy differences between several conformers of the crown ether 1,4,7,10,13,16-hexaoxacyclooctadecane (18-crown-6). Inexpensive force field MMFF94 approximation groups with contemporary semiempirical methods in reproducing the correct order of conformational energies and spatial structures, although the performance in predicting absolute conformational energies compares to standard semiempirical methods. Based on these findings, we suggest a two-step strategy for reliable yet feasible conformational search and sampling in realistic-size flexible organic molecules: i) geometry optimization/preselection of relevant conformers using the MMFF94 force field; ii) single-point energy evaluations using a contemporary semiempirical method. We expect that developed database 37conf8 is going to be useful for development of semiempirical methods.

Introduction

Conformational search and sampling, i.e. selection of a representative set of energetically low-lying structures are the critical steps in the molecular modeling activity involving organic molecules or fragments.[1] Rapid (exponential) growth of the number of conformers with the number of rotatable bonds in the organic moiety makes systematic conformational search with accurate quantum chemistry methods either infeasible or impractical for all but very small systems with 5-6 rotatable bonds. There are two main steps employed to overcome these issues

[a]   Dr. Dmitry I. Sharapa
        Institute of Catalysis Research and Technology, Karlsruhe Institute of Technology (KIT), Hermann-von-Helmholtz Platz 1, Eggenstein-Leopoldshafen, D-76344, Germany
[b]   Dr. Alexander Genaev
        Vorozhtsov Novosibirsk Institute of Organic Chemistry Academician Lavrent'ev Ave., 9, Novosibirsk 630090, Russian Federation
[c]   Prof. Dr. Luigi Cavallo
        KAUST Catalysis Center (KCC), King Abdullah University of Science and Technology, Thuwal-23955-6900, Saudi Arabia. Email: Luigi.Cavallo@kaust.edusa
[d]   Dr. Yury Minenkov
        Moscow Institute of Physics and Technology, Institutskiy Pereulok 9, Dolgoprudny, Moscow Region 141700, Russia
        E-mail: Yury.Minenkov@phystech.edu

and enable practical conformational searches. First, the conformational space is reduced (sometimes drastically) to avoid considering chemically irrelevant energetically high-lying structures via a variety of algorithms, such as vibrational low-mode search, molecular dynamics, stochastic simulations, etc.[2-4] Second, the energy/structure of each generated conformer is calculated with accurate, still computationally inexpensive method. These two steps are tightly connected with each other as application of faster-to-evaluate potential energy function results in larger part of the conformational space covered in fixed time leading to more reliable conformational search/sampling. Hence, an accuracy of computational chemistry methods in relative conformational energies is of paramount importance.

Due to the reasons of efficiency, reliable sophisticated wave function theory (WFT) methods, such as the "gold" standard CCSD(T) method, are restricted to the conformational search for very small systems of dozen of atoms and a few rotatable bonds.[5-13] Application of computationally more affordable (yet accurate) density functional theory (DFT) methods is far more common in conformational search, and it allows to expand the size of the system to a few dozens of heavy atoms and 5-6 rotatable bonds.[14] However, for large realistic-size molecules with more than dozen of rotatable bonds, molecular mechanics (MM) force field (FF) and quantum chemistry semiempirical methods are the only viable options.[15, 16] Apart from so-called "universal" force fields (e.g. UFF[17]) which are of limited accuracy,[18-20] computationally efficient[21] FF methods based on Newton classical physics imply tedious preparation of input files, manual assignment of atom types and require time-consuming parameterization for every new atom type and/or charge/spin-state hampering their application for routine conformational searches. In contrast, semiempirical methods are entirely black box and independent of atom types and hence, are far more easy to use. Moreover, being the cheapest quantum chemistry approach, semiempirical methods provide direct access to the wave function and the properties derived from it, and can straightforwardly handle the systems with different charge and/or multiplicities without reparametrization. It has to be noted that these advantages of semiempirical over FF methods come at increased computational cost. However, with the increasing availability of computational resources, force field methods are being slowly replaced by contemporary semiempirical methods for the conformational search in middle-size systems.

The knowledge that errors in the conformational energies from semiempirical and force field methods can be relevant is based on a limited number of studies.[22],[23],[24],[25] However, these works were benchmarking semiempirical and molecular mechanics energies versus DFT based energies, at best. Considering that the accuracy of DFT methods is known to be limited,[10, 26-28] and that different functionals were used in different studies, a systematic and reliable quantification of the accuracy of semiempirical methods versus accurate energies is missing. In this respect, a valuable contribution was recently published by Rulíšek and co-workers[14] that tested the performance of a set of

contemporary semiempirical methods (PM6*[29-31], PM7[32], DFTB*[33], GFN-xTB[34], OM2*, OM3[35, 36] and RM1[37]) and a number of DFT functionals in reproducing very accurate DLPNO-CCSD(T) conformational energies of smaller peptides and medium-sized macrocycles, or a total of 13 compounds, providing the first reliable estimate of the accuracy of semiempirical methods, although on a limited class of molecules. In this work we expanded the work of Rulíšek and co-workers[14] with a systematic comparison of conformational energies of a variety organic molecules obtained from a list of semiempirical methods with the DLPNO-CCSD(T) counterparts. Differently from Rulíšek and co-workers[14] that focused on a well-defined of 13 peptides, we decided to explore a substantially larger (37 compounds) and chemically very diverse dataset of molecules containing halogens, hydroxyls, ether-, carbonyl-, carboxyl-, amino-, nitro- and other common groups, offering the first reliable calibration of semiempirical methods, including the most recent implementations. Moreover, along with the most recent semiempirical methods we also include standard NDDO schemes and the very accurate[20, 38] and commercially used MMFF94 force field,[39-41] which are still actively used for the conformational search/sampling without any justification of their reliability.

1. Methodology and Computational Details

37Conf8 Database of Accurate Conformation Energies of Organic Molecules

To assess the performance of the semiempirical and FF methods we created a new database of accurate conformational energies and geometries of flexible organic molecules.

Selection Criteria of Molecules to be Included in the Dataset

To become a part of the dataset each molecule was ensured to:
a)   have at least eight unique conformers that are evenly distributed in the range of at least 6 kcal/mol. Narrower interval (1-2 kcal/mol) would result in less reliable statistics (see below);
b)   be free of transition metals and/or any other poorly parameterized elements as Se, Te, etc. According to our recent study[42] blind application of semiempirical methods might result in non-reliable relative conformational energies and/or unrealistic geometries of transition metals;
c)   contain not more than sixty atoms. We introduced this restriction to be able to perform single-point energy reference DLPNO-CCSD(T) calculations with triple-ζ correlation consistent basis set and accurate TightPNO settings in reasonably short time (see below);
d)   to be of practical scientific/industrial interest. All molecules considered in the current work belong to at least one of the following classes: pharmaceuticals, drugs, catalysts, synthetic precursors, industry-related chemicals;
e)   enrich the dataset in chemical diversity. Included systems contain halogens, hydroxyls, ether-, carbonyl-, carboxyl-, amino-, nitro- and other common groups.

Pre-selection of the Conformers

Initially, almost fifty potential candidates were selected based on criteria b–e above. Further, the initial set of structures was pruned to meet criterion a) via the following protocol. First, an extensive conformational search was performed with conformational search routine "scan" using the Tinker[43] package and MMFF94 force field.[39-41] Input files for Tinker were prepared using Open Babel[44] and SDF2XYZ2SDF[45] software. All structures for which less than 8 unique conformers were detected by Tinker or too narrow energy interval was obtained according to MMFF94 force field energy, have been ruled out from the dataset. Second, as FF- and semiempirical/DFT-based conformational potential energy surfaces (PES) can be significantly different, resulting even in different numbers of unique conformers,[46] for each compound survived after previous step 15 structurally/energetically different conformers were manually selected within 10 kcal/mol from the most stable structure found. Third, all 15 conformers for each molecule have been subjected to cheap DFT optimization with PBE functional[47, 48] and λ1 basis sets[49] as implemented in Priroda code[50]. Finally, obtained structures were sorted in the way to keep only 8 conformers providing maximum energetic and geometric diversity. Structural formulas with common names are shown in Figure 1. Few compounds included in dataset do not have a common name or have multiple different names, in this cases names were adjusted and explained in caption.

Protocol to Derive the Reference Geometries and Conformational Energies

After the pre-screening described in Section 2.1.2 we assembled a dataset of 37 compounds with 8 unique DFT-optimized conformers for each. Inspired by S66x8 dataset of Hobza[51, 52] we named it **37Conf8**. The reference conformational geometries and relative energies have been obtained via the following procedure. First, each conformer was reoptimized with the GGA PBE functional as implemented in Gaussian 09[53] suite of programs, as this method was found to accurately reproduce the molecular spatial structures.[54] The Grimme's D3(BJ)[55, 56] dispersion correction was applied to arrive at the PBE-D3(BJ) functional, to account the possible influence of the non-covalent interaction not covered by standard PBE functional on molecular geometries.[57] The default values were adopted for the self-consistent-field (SCF) and geometry optimization convergence criteria. Numerical integration of the exchange-correlation (XC) terms was performed using tighter-than-default "ultrafine" option (pruned, 99 radial shells and 590 angular points per shell) to eliminate the potential numerical noise in energy second derivatives. Geometries were characterized as true energy minima by the eigenvalues of the analytically calculated Hessian matrix.

The all-electron triple-ζ basis sets accomplished with single sets of polarization functions ("def2-tzvp") of Ahlrichs et al.[58] were used on all the elements. The density fitting algorithms with automatic generation of the auxiliary basis sets were turned on by "Auto" Gaussian 09 keyword to speed up the calculations.

The DLPNO-CCSD(T) method of Neese and co-workers[59-68] as implemented in ORCA 4 suite of programs[69, 70] was applied for single-point (SP) energy evaluations on PBE-D3 optimized geometries. This method was shown to provide results of canonical coupled-cluster quality. [71-77] Tighter than the default

"TightPNO" DLPNO settings (TCutPairs = 10^{-5}, TCutPNO = 10^{-7}, and TCutMKN = 10^{-3}) were used, as recommended for applications targeting the most accurate vales.[61, 75, 77] The following triple-ζ correlation consistent basis sets augmented with diffuse functions were used in the present work. Dunning's cc-pVTZ basis sets 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 "packed" conformers.[81-83] However, we do not expect any drastic changes upon going to the DLPNO-CCSD(T)/CBS values 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 "PRECISE" and all geometry optimizations were running with the keyword "DDMAX=0.1" to prevent abrupt jumps in geometry. 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: "theory=qm_n3, tolerance=1e^{-6}". Third, a couple of density 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 code (with "-opt vtight" option). Finally, for the purpose of comparison we included drug design industry appreciated MMFF94 force field method as implemented in TINKER code[43] (with "newton" routine and "RMS Gradient 0.001").
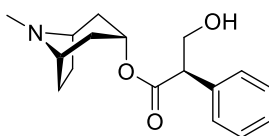
The methods have been tested in reproducing relative conformational energies via SP energy evaluations on geometries obtained in Section 2.2.3 and geometry optimizations thereof.
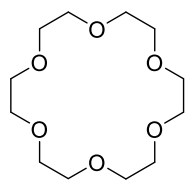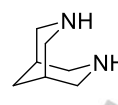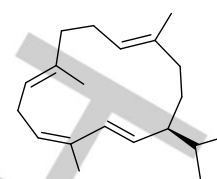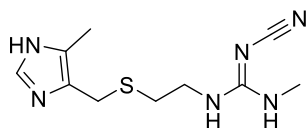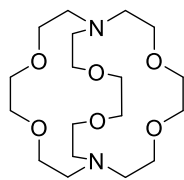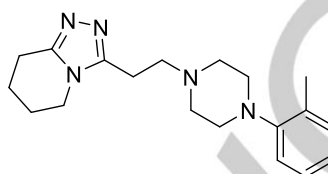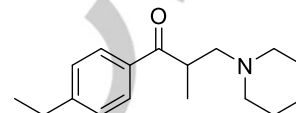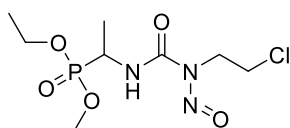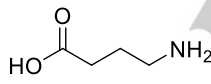
18-crown-6
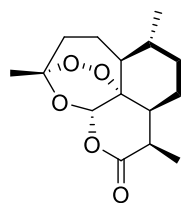
Artemisinin
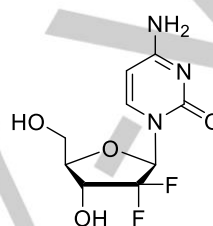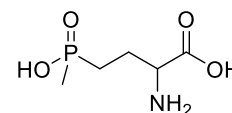
Atropine

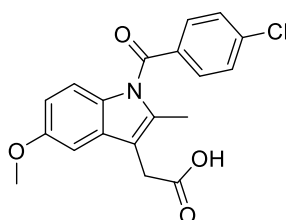Bispidine

Cembrene

Cimetidine

Cryptand-222

Dapiprazole

Eperisone

Fotemustin

GABA

Gemcitabine

Glufosinate

Indometacin
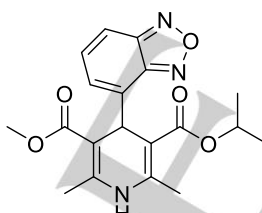
Isradipine
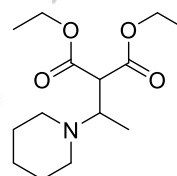
Knoevenagel_cond [a]
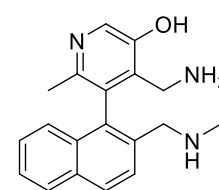
Liu_catal [b]
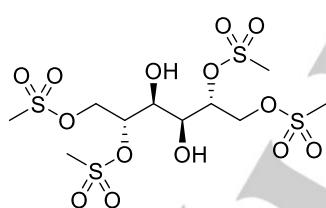
Mannosulfan
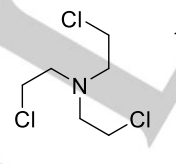
Nitrogen mustard
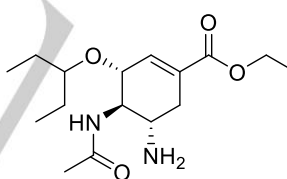
Oseltamivir
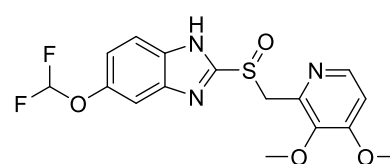
Pantoprazole
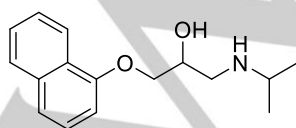
Propranolol
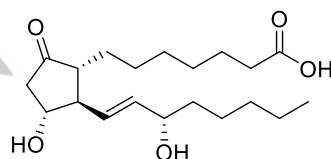
ProstaglandinE1
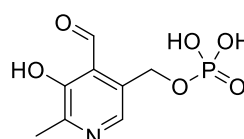
Pyridoxal phosphate
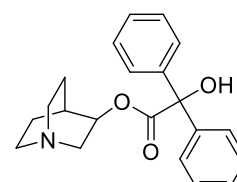
Quinuclidinyl benzilate
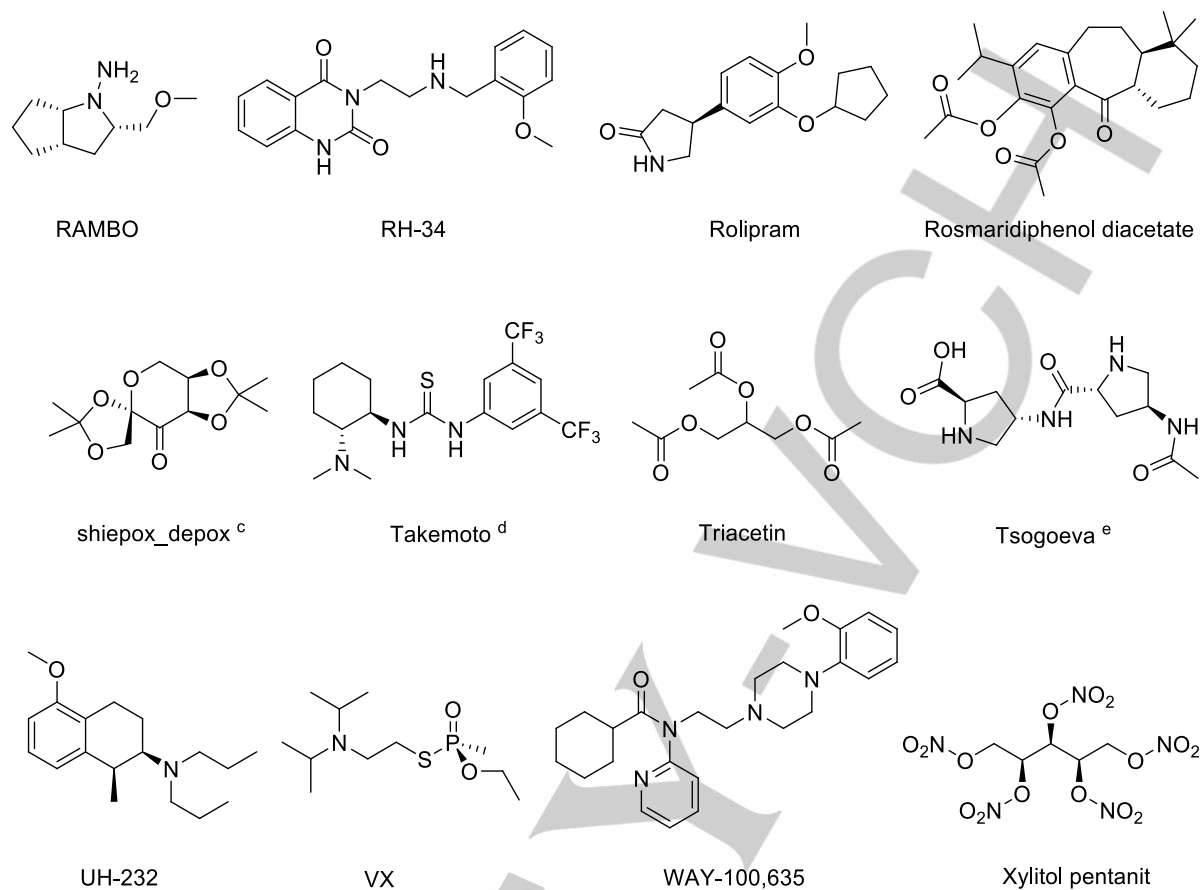
Figure 1. 37Conf8 dataset. a) Intermediate in catalytic cycle of Knoevenagel Condensation, where piperidine plays role of organocatalyst; b) axially chiral pyridoxamines used (as catalyst) by Liu et al in asymmetric enantioselective transamination of α-keto acids[93] c) catalyst used in Shi epoxidation[94] 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 (ρ) via the following formula:

$$\rho(X, Compound) = \frac{\sum_{i=1}^{n}(E_{x,i} - \overline{E}_x)(E_{y,i} - \overline{E}_y)}{\sqrt{\sum_{i=1}^{n}(E_{x,i} - \overline{E}_x)^2 \sum_{i=1}^{n}(E_{y,i} - \overline{E}_y)^2}}$$

(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 is the conformational energy of ith conformer, and $\overline{E}$ is the average conformational energy for n conformers from a given method. The ρ coefficient can occupy any value in the interval [-1, +1]. If the ρ value is close to 1 there is an absolute correlation, and if this value is close to -1 there is an anticorrelation.

While the ρ^2 is the most popular criterion in the field of chemoinformatics to quantitatively measure the correlation (ρ^2 ≥0.95 indicates excellent correlation), it can be misleading when 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) ρ^2 values can correspond to both correlation ($\rho \approx 1$) and anti-correlation ($\rho \approx -1$).

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:

$$MAE(X, Compound) = \frac{\sum_{i=0}^{n}|E_i(X) - E_i(Y)|}{n}$$

(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 "long" or labile molecules: minor change in a single torsion/dihedral angle in the middle of the molecule results in large RMSD for the atoms located at the "tails" of the system. 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.

$$MAE(X, Compound) = \frac{1}{mn}\sum_{i=1}^{n}\sum_{j=1}^{m}|D_{ij}(X) - D_{ij}(Y)|$$

(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 was utilized as a measure of method's 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 ρ=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 ρ and quartile 1 (Q1) – quartile 3 (Q3) span.

The first group of methods demonstrating rather moderate performance is formed by "standard" NDDO-based schemes. The methods within this group appear in the order of their historical development. The MNDO scheme (1977) shows the worst performance with ρ=0.52 and Q1/Q3 values of -0.10/0.79. Somewhat better results, namely ρ of 0.67 and Q1/Q3 of 0.15/0.85, were obtained for AM1 (1985) method. Finally, even better ρ 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 ρ of 0.73 and Q1/Q3 of 0.46/0.84 which is slightly better comparing to its competitor AM1. PM6 (2007) method with ρ=0.61 and Q1/Q3

values of 0.29/0.90 demonstrated slightly worse performance comparing to its predessesor 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 ρ=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 ρ values ranging from -0.95 (MNDO) to -0.29 (PM7) have been obtained implying anticorrelation, we believe all "standard" NDDO-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 conformational energies in these cases are unrealistic. Our results on the performance of semiempirical methods are more in line with the recent paper of Řezáč et al.,[14] even though it was based on significantly smaller dataset both in size and diversity (MPCONF196 data set contains of 196 conformation of 13 small peptides or macrocycles). Similarly to our conclusions, Řezáč and co-workers[14] found PM7 not being reliable for conformational sampling, and claim that other modern techniques have to be taken with a good portion of precaution. 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 classified as "good" and consists from recently developed and 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 ρ 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 significantly "standard" semiempirical methods which is 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 taking 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.
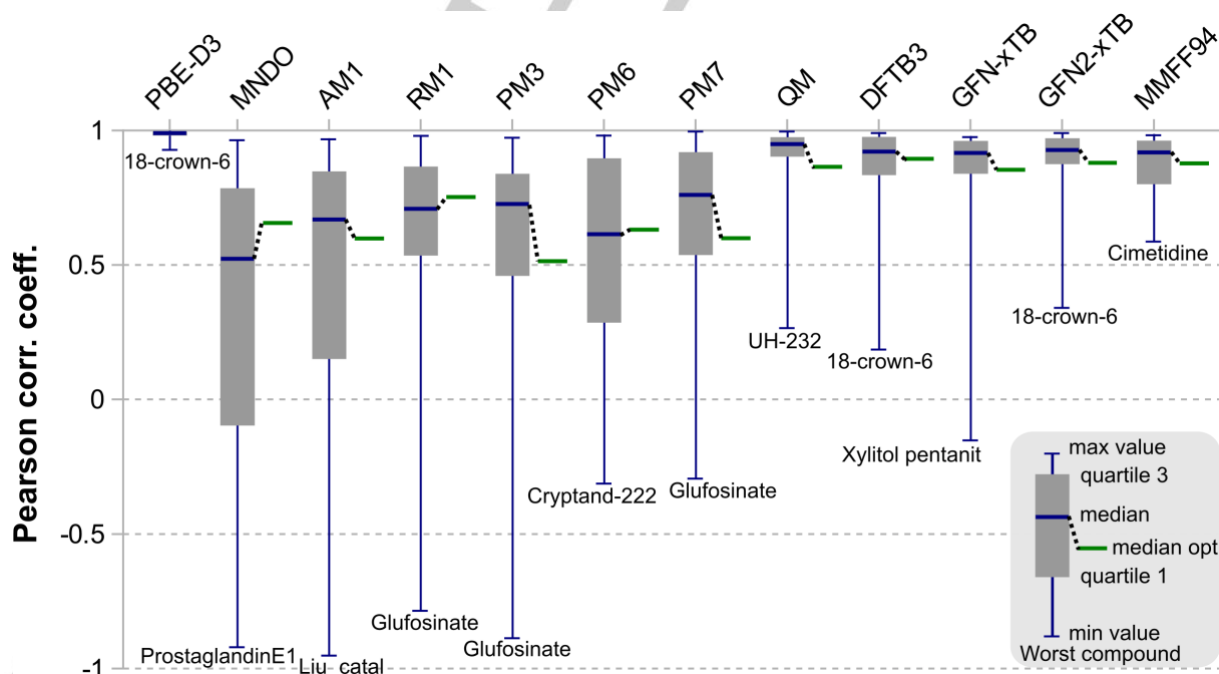
Figure 2. Pearson correlation coefficients (medians, quartile 1 (Q1), quartile 3 (Q3), highest and smallest ρ values) obtained for correlation between Single-Point (SP) energy based DFT, semiempirical and force field conformational energies and their reference DLPNO-CCSD(T)/cc-pVTZ counterparts. Median 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 Person correlation coefficients, the methods clearly form the two separate groups. Again, the first group of methods consists of "standard" 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 "standard" semiempirical methods, substantially smaller Q1/Q3 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 approximation with "standard" semiempirical methods. Thus, 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.
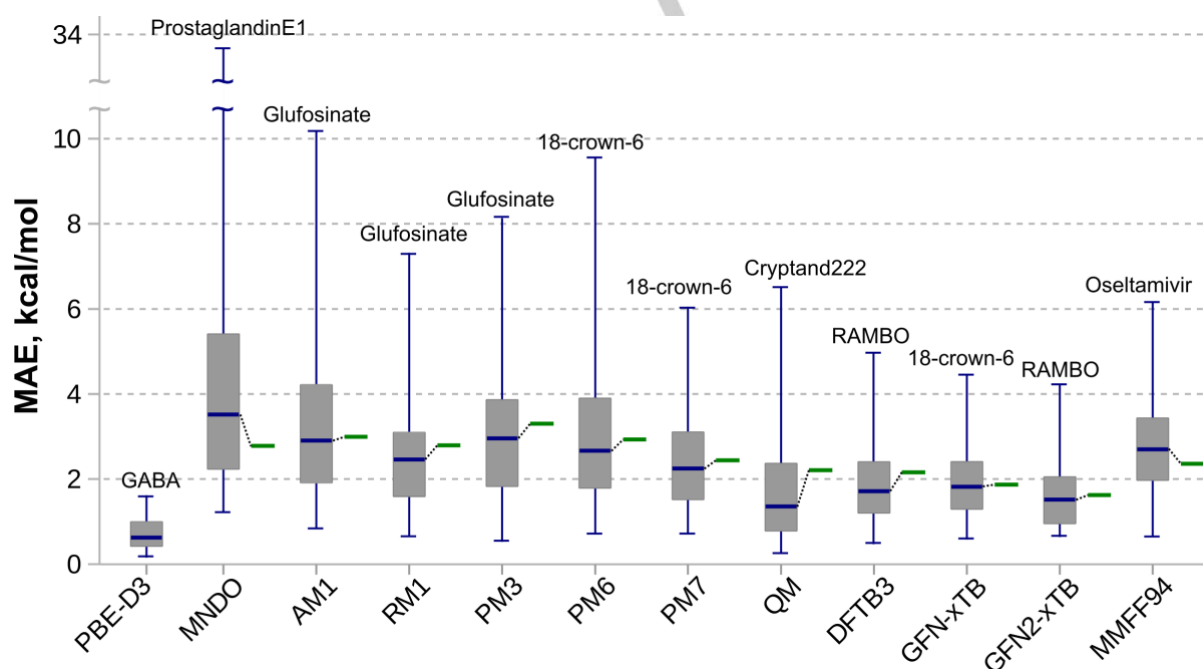
Figure 3. Absolute errors (medians, quartile 1 (Q1), quartile 3 (Q3) and highest/smallest mean absolute error values) obtained for Single-Point (SP) energy based DFT, semiempirical and force field conformational energies with respect to their reference DLPNO-CCSD(T)/cc-pVTZ counterparts. Median values 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_i, 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- symmetrical

conformers (№3 and №7 in **37Conf8** respectively). To illustrate these failures, in Figure 4 the energy difference between energies of $D_{3d,}$ conformers and and D_2/S_6 conformers is depicted.
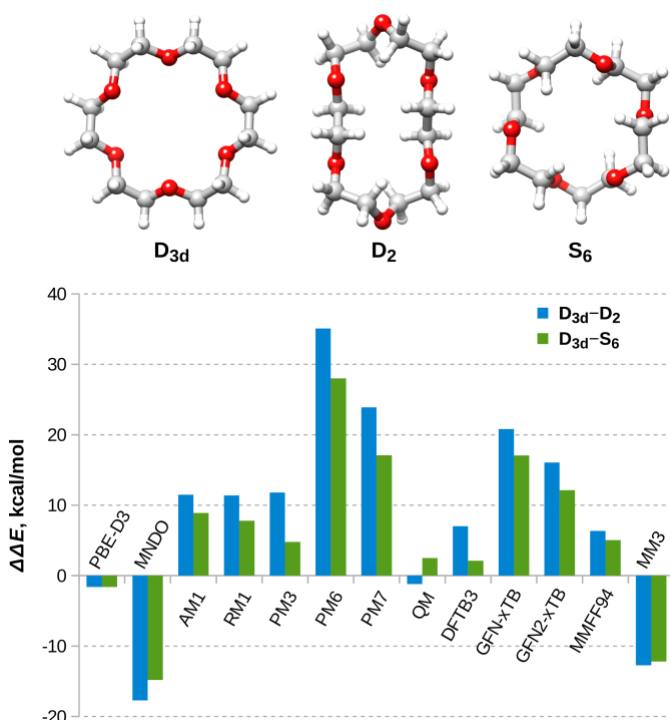
Figure 4. Energy difference

$$\Delta\Delta E = (E_{D_{3d}}^{method} - E_{D_2/S_6}^{method}) - (E_{D_{3d}}^{DLPNO-CCSD(T)} - E_{D_2/S_6}^{DLPNO-CCSD(T)})$$

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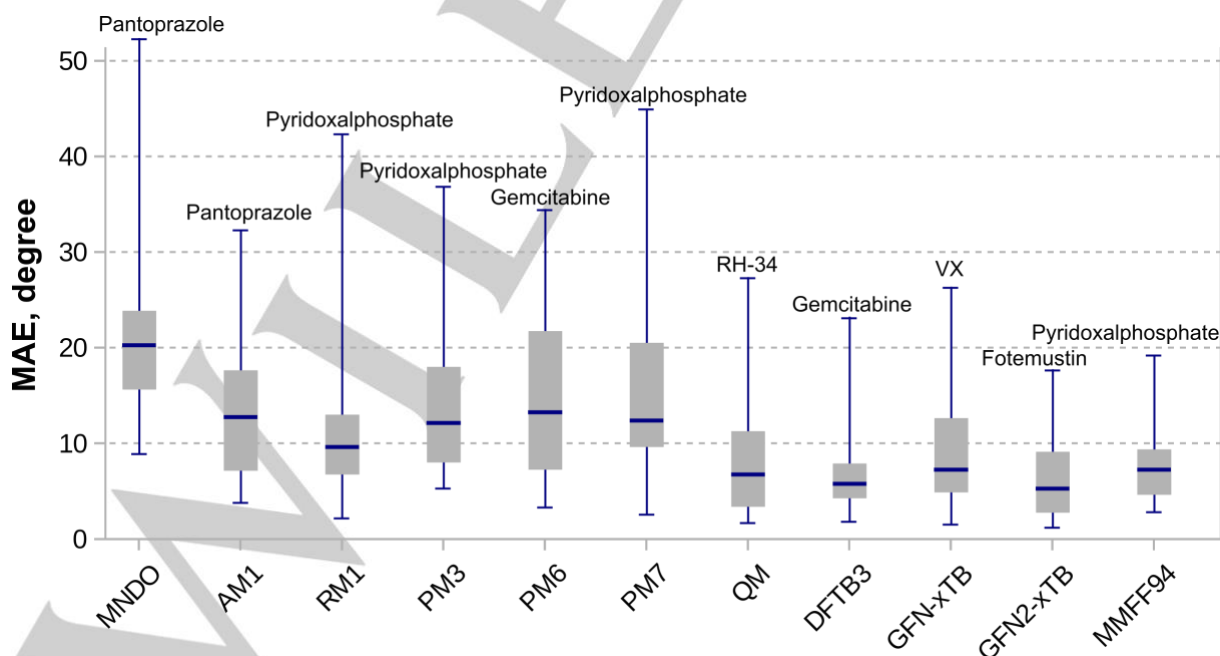
Figure 5. Calculated errors in the values of dihedral angles (degrees) corresponding to rotatable bonds obtained upon geometry optimizations with all but PBE-D3 tested methods.

Similarly to what was found in Section 3.1 the methods for the two groups based on their performance in reproducing PBE-D3 dihedral angles. Thus, "standard" semiempirical methods 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 utilized to test a variety of "standard" MNDO, AM1, RM1, PM3, PM6, PM7 and "contemporary" QM, DFTB, GFN-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

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, 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. 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.^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. Cavasin, 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. Shiri *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.^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. Sivalingam, 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.
[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] H. Valdés, V. Klusák, M. Pitoňák, O. Exner, I. Starý, P. Hobza, L. Rulíšek *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-β-D-erythro-hexo-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.

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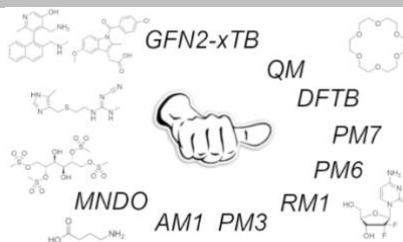
Dmitry I. Sharapa, Alexander Genaev, Luigi Cavallo* and Yury Minenkov*

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