Application of Semiempirical Methods to Transition Metal Complexes: Fast Results but Hard-to-Predict Accuracy.

Yury Minenkov, Dmitry I. Sharapa, and Luigi Cavallo

J. Chem. Theory Comput., Just Accepted Manuscript • DOI: 10.1021/acs.jctc.8b00018 • Publication Date (Web): 22 May 2018

Downloaded from http://pubs.acs.org on May 29, 2018
Application of Semiempirical Methods to Transition Metal Complexes: Fast Results but Hard-to-Predict Accuracy.

Yury Minenkov,*§,† Dmitry I. Sharapa‡ and Luigi Cavallo*†

§ Moscow Institute of Physics and Technology, Institutskiy Pereulok 9, Dolgoprudny, Moscow Region 141700, Russia

†KAUST Catalysis Center (KCC), King Abdullah University of Science and Technology, Thuwal-23955-6900, Saudi Arabia

‡ Chair of Theoretical Chemistry and Interdisciplinary Center for Molecular Materials Friedrich-Alexander-Universität Erlangen-Nürnberg Egerlandstraße3, 91058 Erlangen, Germany


ABSTRACT.

A series of semiempirical PM6* and PM7 methods has been tested in reproducing of relative conformational energies of 27 realistic-size complexes of 16 different transition metals (TM).
An analysis of relative energies derived from single-point energy evaluations on density functional theory (DFT) optimized conformers revealed pronounced deviations between semiempirical and DFT methods, indicating fundamental difference in potential energy surfaces (PES). To identify the origin of the deviation, we compared fully optimized PM7 and respective DFT conformers. For many complexes, differences in PM7 and DFT conformational energies have been confirmed often manifesting themselves in false coordination of some atoms (H, O) to TMs and chemical transformations/distortion of coordination center geometry in PM7 structures. Despite geometry optimization with fixed coordination center geometry leads to some improvements in conformational energies, the resulting accuracy is still too low to recommend explored semiempirical methods for out-of-the-box conformational search/sampling: careful testing is always needed.

1. Introduction

Transition metal (TM) complexes bearing large bulky ligands are examples of important catalysts, both in vivo and in the chemical and pharmaceutical industry.\textsuperscript{1-5} To better understand their catalytic activity and physical properties, theoretical modelling of these complexes is today routinely undertaken and accounts for a significant fraction of the computational resources used worldwide.\textsuperscript{6-11} The complex electronic structure of TM (which is responsible for their catalytic activity), implies a number of computational challenges for an accurate modelling, such as relativistic effects, dynamic and static correlation effects, poor scalability of the reliable theoretical methods, and multiple low-lying electronic states.\textsuperscript{8, 12-20} Numerous strategies have been proposed to overcome these issues, and many instructive results have been obtained.\textsuperscript{21-47}
additional important challenge, the strong conformational flexibility of many TM species, has so far received only limited attention from selected research groups, and is to large extent still an unsolved problem. This leads to known difficulties in the modeling of large TM complexes containing multiple rotatable bonds.

Accurate conformational search and/or sampling have been shown to be relevant for a number of cases related to molecular modelling. First, these are useful for modelling of chemical reactivity since a reaction may not arise from the most stable conformation, and conformational transitions are often required to initiate a chemical reaction. Moreover, as shown by Besora and co-workers, the errors originated from choosing wrong conformations can alter the calculated energy profiles by as much as 10 – 20 kcal/mol. Second, reliable conformational sampling is often required for interpretation of physical experiments, like NMR, gas phase electron diffraction or dipole-moment experiments, in order to perform a proper Boltzmann averaging over the conformational space. Third, conformational sampling is sometimes needed for quantitative structure activity relationship (QSAR) studies aimed at identifying better catalysts or drug design because molecular descriptors can be sensitive to the conformations, and using only the most stable conformation might not be sufficient for reliable predictions.

To solve these issues and make accurate calculations of conformational flexible systems possible, a fast and reliable approach to sample the conformational space of the molecule is of paramount importance. Needless to say, this is a daunting task since the complexity of this problem is exponential \( O^N \). A systematic conformational search thus implies minimization of the energy function ca. \( 3^N \) times, where \( N \) is the number of rotatable bonds. To get reliable relative energy estimates for TM complexes, the energy function from sophisticated wave or density function theory (DFT) should be used. These energy functions, however, are so
computationally intensive that only systems with a few dozens of atoms and $3 - 4$ rotatable bonds can be subjected to an exhaustive conformational search. A possible way to solve these issues is to use computationally inexpensive energy functions.

The cheapest in terms of CPU-time approach to treat the transition metal complexes is represented by the force field (FF) methods. In principle, even the standard and very general out-of-the-box FF methods, such as the universal force field (UFF) approximation, have been shown to perform reasonably well for conformational sampling of transition metal complexes. In cases where the UFF method is not sufficiently accurate, a few more FF-based strategies can be utilized. Thus, careful parameterization of both metal and functional groups can be done prior to the conformational search. However, the FF parameterization is a tedious task requiring expertise and can be time consuming. Therefore, this approach is not suitable for routine applications as in molecular screening, where the conformational sampling of thousands of molecules is required. In any case, parameterization of a FF is probably the most viable option in case of conformational search on a single transition metal complex with an immutable coordination sphere, or when large amount of structures has to be sampled. Another FF-related approach requiring less manual work is to fix the atoms comprising the transition metal coordination sphere either at their crystallographic or DFT positions, and run the conformational search on the organic “non-frozen” part of the complex. Semiempirical methods represent another out-of-the-box alternative to FF methods, albeit they are computationally more intensive, and have been successfully applied for decades in many areas of organic and to less extent transition metal chemistry. Thus, PM3(TM) and PM6 semiempirical methods have been successfully used for conformational search on molybdenum, ruthenium, rhenium and technetium systems. However, their accuracy...
for relative conformational energies have never been systematically tested for a series systems containing various transition metals. In addition, there is no single opinion on the preference of FF or semiempirical methods for the conformational search of transition metal species. Comparing FF and semiempirical methods to sample the conformational spaces of a series of technetium complexes, Buda et al. noted that in a number of cases semiempirical methods are more reliable than FFs, as the latter could not predict the right isomer. In contrast, Gillespie et al. showed that a few potential energy minima found during molecular mechanics conformational search of rhenium complex have disappeared after re-optimization with semiempirical methods. Taken together these studies indicate that the conformational energy surface of particular complex apparently depends (sometimes strongly) on the particular method chosen for geometry optimization, hampering the reliable conformational search.

In the current work we seek to contribute to testing an accuracy of semiempirical methods and generate a database containing 10 very accurate relative conformational energies for each out of 27 realistic sized (ca. 100 atoms) TM complexes initially retrieved from the Cambridge Structural Database. The developed database was used to test relative conformational energies from widely available out-of-the-box semiempirical methods: PM6, PM6-D3, PM6-DH+, PM6-DH2, PM6-DH2X, PM6-D3H4, PM6-D3H4X, and PM7. The performance of all methods was carefully analyzed, the most difficult cases were identified and the origin of the failures discussed in detail.

2. Computational Details

2.1 Selection of Complexes and Conformer Generation
All the initial structures of the transition metal complexes studied in the present work have been retrieved from the Cambridge Structural Database (CSD). Pursuing the maximal diversity in our benchmark set we included complexes of 16 transition metals related to catalysis (Au, Cd, Cr, Hf, Hg, Ir, Pd, Pt, Rh, Ru, Sc, Ti, V, W, Y, Zr) containing various functional groups and different covalent and non-covalent ligands. For each complex we have generated 10 conformers in total from the X-Ray structures as follows. First, the random values of the dihedral angles corresponding to the rotatable bonds had been set up and the preliminary structure was generated. Then, to avoid clashing, each structure was pre-optimized with UFF potential by keeping all the bonds and angle bends at their X-Ray values and relaxing only the dihedral angle values of the rotatable bonds. All the transition metal complexes studied in this work together with their CSD names are given in Figure 1.
Figure 1. Transition metal complexes studied in this work together with their Cambridge Structural Database (CSD) names.

2.2 Technical Protocols Utilized to Derive and Compare the Conformational Energies

2.2.1 Strategy I

In the first part of the study all geometry optimizations of the generated conformers (see Section 2.1) were performed with the local GGA PBE\textsuperscript{125-126} functional as implemented in Gaussian 09\textsuperscript{127} suite of programs as this method was found to accurately reproduce the molecular spatial structures of inorganic species.\textsuperscript{34, 128-132} The Grimme’s D3(BJ)\textsuperscript{118} 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.\textsuperscript{133-135} 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 double-ζ basis sets accomplished with single sets of polarization functions (“def2-svp”) of Ahlrichs et al.\textsuperscript{136} were used on all the elements with $Z \leq 36$. On all the elements with $Z > 36$ to account for the scalar relativistic effects the Stuttgart-type effective core potentials\textsuperscript{137} were used to describe the inner (28 electrons on Sn, Cd, Pd, I, Te, Rh, Ru, Y and Zr, 60 electrons on Au, Hf, Hg, Ir, W and Pt) electrons in combination with corresponding def2-svp
basis set.\textsuperscript{136} 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 single-point energy evaluations on the structures obtained after geometry optimizations were performed with the hybrid-meta-GGA M06 functional as implemented in Gaussian 09. Alike in the geometry optimization procedure, numerical integration of the exchange-correlation (XC) terms was performed using “ultrafine” grid to eliminate the potential numerical noise in electronic energy.\textsuperscript{138-139} The all-electron triple-ζ basis sets accomplished with single set of polarization functions (“def2-tzvp”) of Ahlrichs et al.\textsuperscript{136} were used on all the elements with $Z \leq 36$. On all the elements with $Z > 36$ to account for the scalar relativistic effects the Stuttgart-type effective core potentials\textsuperscript{137} were used to describe the inner electrons (28 electrons on Sn, Cd, Pd, I, Te, Rh, Ru, Y and Zr, 60 electrons on Au, Hf, Hg, Ir, Os and Pt) in combination with corresponding def2-tzvp basis set.\textsuperscript{136}

Thus obtained conformational energies have been used as references and are denoted as M06/def2-tzvp/PBE-D3/def2-svp. To show that thus obtained relative conformational energies are not sensitive to the choice of particular DFT method, we have performed a series of additional SP energy calculations with PBE0-D3 method and def2-tzvp basis set abbreviated further as PBE0-D3/def2-tzvp/PBE-D3/def2-svp. In addition, to explore the basis set effect, the M06/def2-tzvp/PBE-D3/def2-svp relative conformational energies have been systematically compared with those derived from geometry optimization method, PBE-D3/def2-svp. Notably, practically in all cases the PBE0-D3/def2-tzvp/PBE-D3/def2-svp and PBE-D3/def2-svp conformational energies were found to practically mimic their M06/def2-tzvp/PBE-D3/def2-svp counterparts.
Finally, the SP energies on the optimized PBE-D3 conformers were calculated with out-of-the-box semiempirical PM6, PM6-D3, PM6-DH+, PM6-DH2, PM6-DH2X, PM6-D3H4, PM6-D3H4X and PM7 methods as implemented in MOPAC electronic structure package.  

2.2.2 Strategy II

In the second part of the work each structure generated in Section 2.1 and optimized with PBE-D3/def2-svp protocol as outlined in Section 2.2.1 has been re-optimized with PM7 semiempirical method as implemented in MOPAC electronic structure package. All the default MOPAC internal values were used apart from the geometry optimization termination criterion GNORM, for which the default value of 1 kcal·mol⁻¹/Å was decreased to 0.1 kcal·mol⁻¹/Å to ensure more precisely optimized structures. The two geometry optimization strategies were used for optimizing the conformers. In the first strategy, the “non-constrained” geometry optimization has been performed with the positions of all the atoms relaxed during the optimization process. In the second strategy, usual for routine conformational search in the transition metal catalysis, the “constrained” geometry optimization has been performed with fixed (at their PBE-D3/def2-svp values) positions of transition metal and bonded to transition metal atoms to preserve the structure of the coordination center. In addition, the positions (Cartesian coordinates) of few more atoms constituting several functional groups directly bonded to transition metal were kept fixed in some complexes at their DFT positions as well: GeCl₃ group in TITVEX01 (Ph₃P)₃AuGeCl₃, CSe group in DEFVIT [(MeO)₃P]₃Cr(CO)₂(CSe), and CO group in CAFKOJ [(MeO)₃P]₃Ir(CO), DEFVIT, FEGGII (-CH₂-)C(CH₂OBn)₂-(CH=CH₂)-Ir(PPh₃)₂(CO).
In both PM7 optimization strategies after the PM7 relative energies for each conformers were derived and tabulated, each conformer was again re-optimized with PBE-D3 method (as in Section 2.2.1) and M06 and PBE0-D3 SP energies with def2-tzvp basis sets were calculated to derive the relative conformational energies as it was done in Section 2.2.1.

2.3 Comparing the Conformational Energies

Few quantitative criteria have been used to compare the semiempirical conformational energies with their reference values.

2.3.1 Calculation of Pearson Correlation Coefficient

To quantify the correlation between conformational energies from semiempirical computational chemistry methods, PBE0-D3/def2-tzvp//PBE-D3/def2-svp and PBE-D3/def2-svp protocols with reference M06/def2-tzvp//PBE-D3/def2-svp conformational energies, we have calculated the Pearson correlation coefficient ($\rho$) and the squared version thereof ($\rho^2$).

The following formula was used to calculate the Pearson correlation coefficient:

$$\rho(X, Y) = \frac{\sum_{i=1}^{n}(E_{X,i} - \bar{E}_X)(E_{Y,i} - \bar{E}_Y)}{\sqrt{\sum_{i=1}^{n}(E_{X,i} - \bar{E}_X)^2 \sum_{i=1}^{n}(E_{Y,i} - \bar{E}_Y)^2}}$$  (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. M06/def2-tzvp//PBE0-D3/def2-svp, $n$ is the number of conformations calculated for a given transition metal complex, $E_i$ is the conformational energy of $i^{th}$ conformer, and $\bar{E}$ is the average conformational energy for $n$ conformers from a given method. The $\rho$ 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 anti-correlation.

The squared Pearson correlation coefficient (ρ²) is obtained straightforwardly from Eq. 1 and can be any value in the interval [0, 1]. While the ρ² is the most popular criterion in the field of chemoinformatics to quantitatively measure the correlation (ρ² ≥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) ρ² values can correspond to both correlation (ρ ≈ 1) and anti-correlation (ρ ≈ -1).

2.3.2 Calculation of the Mean Unsigned Deviations in Conformational Energies

Another criterion to judge the quality of the relative conformational energies is to calculate the mean absolute deviation (MUD) between the relative energies obtained with particular method and corresponding reference values (M06/def-tzvp//PBE-D3/def2-svp). The following formula was used to calculate the mean absolute deviation of particular complex:

\[ MUD(\text{Complex}) = \frac{\sum_{i=1}^{n} |E_i(X) - E_i(Y)|}{n} \]  \hspace{1cm} (2)

Where X is the tested method, Y is the reference method (M06/def-tzvp//PBE-D3/def2-svp), n is the number of conformers calculated for the certain complex and E_i is the relative energy of the conformer i.

3. Results and Discussion
The results and discussion is organized as follows. First, following the Strategy I, we explored the potential energy surface (PES) from several electronic structure theory methods. For each structure/conformer in our dataset we compared the relative conformational energies retrieved from the semiempirical, PBE-D3/def2-svp, PBE0-D3/def2-tzvp and M06/def2-tzvp SP energies on the PBE-D3 optimized geometries. Second, following the Strategy II, we estimated the potential errors arising from the conformational search based on geometry optimization with semiempirical methods. We performed geometry optimizations with PM7 method of all the conformers generated in the previous step followed by subsequent PBE-D3/def2-svp geometry optimizations and PBE0-D3/def2-tzvp and M06/def2-tzvp SP energy evaluations. Finally, the results are analyzed, the possible origins of the largest deviations are discussed, and conclusions on the applicability of the semiempirical methods for the conformational search and sampling in the realistic-size transition metal complexes are given.

3.1 Comparing the Relative Conformational Energies Using Strategy I

The averaged over all the complexes $\rho^2$ and $\rho$ (Pearson correlation coefficient) values of conformational energies derived from the SP energies obtained with all the methods tested on the PBE-D3/def2-svp geometries with respect to the reference M06/def2-tzvp values are given in Figures 2 and 3. In addition, the lowest and highest $\rho^2$ and $\rho$ values are also given in the Figures through the solid lines to have an idea of the span.

As can be seen from Figure 2, an average PBE0-D3/def2-tzvp and PBE-D3/def2-svp $\rho^2$ values turned out to be 0.97 and 0.93, indicating rather strong correlation. Quite large average $\rho^2$ numbers and an absence of any negative $\rho$ values for any complex studied in the present work indicate DFT methods to be quite robust in predicting the relative conformational energies. It
suggests the M06/def2-tzvp//PBE-D3/def2-svp protocol to be the reliable source of relative
conformational energies, and we do not expect any significant changes upon change in the
reference SP and geometry optimization method with any other contemporary DFT method
accomplished with the basis set of reasonable quality.

Remarkably, for all tested semiempirical methods significantly lower average $\rho^2$ values all below
0.55 have been obtained. The lowest average values of 0.45 and 0.47 were obtained for PM7 and
PM6 methods, correspondingly. The highest average value of 0.55 were obtained for PM6-D3H4
and PM6-D3H4X methods, indicating only insignificant difference between the best and the
worth performers. The performance of the semiempirical methods is heterogeneous and varies
from complex to complex which is reflected in the large span between the largest and lowest $\rho^2$
values. Thus, for the most recent PM7 method the smallest $\rho^2$ value of only 0.02 has been
obtained for DEFVIT complex $[(\text{MeO})_3\text{P}]_3\text{Cr(CO)}(\text{CSe})$ indicating practically no correlation.
Meanwhile, quite high $\rho^2$ value of 0.93 was obtained for complexes FONLID $\text{Hg}[\text{CH(COOEt)}_2]_2$
and DUGVEH $[(\text{Me}_3\text{SiCH}_2)_3\text{Te}]_3\text{RhCl}_3$ which is comparable to DFT performance.

Another striking difference comparing to the tested DFT methods is existing of negative $\rho$ values
which, according to Figure 3, have been detected for all the semiempirical methods indicating
signs of anti-correlation in some cases. For example, for PM7 method negative $\rho$ value was
obtained for 6 out of 27 complexes: TITVEX01 $(\text{Ph}_3\text{P})_3\text{AuGeCl}_3$ ($\rho=-0.48$), CAFKOJ
$[(\text{MeO})_3\text{P}]_3\text{Ir(CO)}$ ($\rho=-0.42$), AKUGOD $[(\text{Cy})_2\text{P-OH}]_2\text{PtCl}_2$ ($\rho=-0.42$), BOBXAS
$(\text{Bn})_3\text{Sc(Furan)}_3$ ($\rho=-0.48$), GENQIZ $\text{Ti(NPh)}_2$4 ($\rho=-0.40$) and MOGWIP $(\text{Bn})_3\text{Y(Furan)}_3$ ($\rho=-0.18$).
Figure 2. The $\rho^2$ (the squared Pearson correlation coefficient) values obtained for correlation between single-point (SP) energy-based DFT (PBE0-D3/def2-tzvp and PBE-D3/def2-svp) and semiempirical (PM6, PM6-D3, PM6-DH+, PM6-DH2, PM6-DH2X, PM6-D3H4, PM6-D3H4X and PM7) relative conformational energies and their corresponding references (M06/def2-tzvp) for transition metal complexes studied in the present work. The solid bars indicate the average values, and the ends of the solid lines at each bar give the lowest and the highest values.
Figure 3. The $\rho$ (the Pearson correlation coefficient) values obtained for correlation between single-point (SP) energy-based DFT (PBE0-D3/def2-tzvp and PBE-D3/def2-svp) and semiempirical (PM6, PM6-D3, PM6-DH+, PM6-DH2, PM6-DH2X, PM6-D3H4, PM6-D3H4X and PM7) relative conformational energies and their corresponding references (M06/def2-tzvp) for transition metal complexes studied in the present work. The solid bars indicate the average values, and the ends of the solid lines at each bar give the lowest and the highest values.

To explore the quality of the absolute values of the relative conformational energies important for reliable Boltzmann distribution, for each complex we calculated the mean absolute deviations (MUDs) between conformational energies obtained with tested method and their reference (M06/def2-tzvp) counterparts. The results are summarized in Figure 4.
Figure 4. The mean unsigned deviation (MUD) values obtained between reference M06/def2-tzvp reference conformational energies and their PBE0-D3/def2-tzvp, PBE-D3/def2-svp, semiempirical (PM6, PM6-D3, PM6-DH+, PM6-DH2, PM6-DH2X, PM6-D3H4, PM6-D3H4X and PM7) counterparts for transition metal complexes studied in the present work. The solid bars indicate the average MUD values, and the ends of the solid lines at each bar give the lowest and the highest MUD values obtained for particular complexes.

For the PBE0-D3/def2-tzvp method the average MUD turned out to be 0.6 kcal/mol. The PBE-D3/def2-svp method with an average MUD of 1.0 kcal/mol is only slightly less accurate. Significantly larger MUDs in range 3.1 (PM6-D3H4 and PM6-D3H4X) to 3.8 (PM7) kcal/mol have been obtained for semiempirical methods. Comparing to DFT methods, notably larger variations in MUD values have been obtained from complex to complex. Thus, the largest PM7 MUD of 11.5 kcal/mol was obtained for CAFKOJ complex [(MeO)₃P]₃Ir(CO) and the lowest MUD of 0.6 kcal/mol was obtained for FONLID complex Hg[CH(COOEt)₂]₂. Combined with respective ρ² and ρ tests, the results indicate quite heterogeneous performance in reproducing of
the reference conformational energies: good performance for some complexes alternates with spectacular failures for others.52

3.2 Comparing the Relative Conformational Energies Using Strategy II

Since the semiempirical methods sometimes fail to reproduce the coordination center of transition metal complexes even qualitatively, we performed the PM7 geometry optimizations in the two ways: with relaxation of all parameters (unconstrained geometry optimization) and with fixed atomic positions of the coordination center atoms, see Computational details section 2.2.2.

3.2.1 Unconstrained Geometry Optimization

The $\rho^2$ and $\rho$ values between PM7//PM7 and M06-def2-tzvp//PBE-D3/def2-svp relative conformational energies are presented in Figure 5. The average $\rho^2$ value obtained for all 27 structures turned out to be 0.41 which is reasonably close to PM7 $\rho^2$ value of 0.45 obtained based exclusively on SP energies. An average $\rho$ value turned out to be 0.25 which is smaller than that of 0.44 based on the SP energies. The largest $\rho^2$ values followed by positive $\rho$ values have been detected for TITVEX01 $(\text{Ph}_3\text{P})_3\text{AuGeCl}_3$ ($\rho^2/\rho=0.85/0.92$), REPFID $(\text{HexylCyt})_2\text{CdBr}_2$ ($\rho^2/\rho=0.80/0.89$), FONLID $\text{Hg}[\text{CH(COOEt)}_2]_2$ ($\rho^2/\rho=0.90/0.95$), DILQAC $[\text{MeOC(OMe)CH(Me)NH-PPh}_2]_2\text{PtCl}_2$ ($\rho^2/\rho=0.83/0.91$) and DUGVEH $[(\text{Me}_3\text{SiCH}_2)_2\text{Te}]_3\text{RhCl}_3$ ($\rho^2/\rho=0.81/0.90$) indicating correlation. In contrast, large $\rho^2$ value followed by negative $\rho$ value was obtained for REMVOW [(propen-2-y1)-CH=C=\text{Ru(PCy}_3)_2\text{Cl}_2$ ($\rho^2/\rho=0.83/-0.91$) indicating the anti-correlation, and making black-box application of PM7 method for this complex dangerous.
Figure 5. The $\rho^2$ and $\rho$ correlation coefficients obtained between PM7 relative conformational energies on PM7 geometries and M06/def2-tzvp relative conformational energies on PBE-D3/def2-svp optimized geometries with PM7 optimized geometries used as starting point.
Figure 6. The mean unsigned deviation (MUD) values obtained between the PM7//PM7 and M06/def2-tzvp//PBE-D3//def2-svp relative conformational energies for transition metal complexes studied in the present work. The solid bars indicate the average MUD value, and the ends of the solid lines at each bar give the lowest and the highest absolute deviation obtained for particular complexes.
In Figure 6 the MUDs obtained between PM7//PM7 and M06/def2-tzvp//PBE-D3/def2-svp conformational energies are presented for each complex together with the largest and lowest absolute deviations in the relative conformational energies. The average MUD turned out to be 6.5 kcal/mol. The lowest MUD of 1.0 kcal/mol has been obtained for complex FONLID Hg[CH(COOEt)2]2 and the largest MUD of 34.9 kcal/mol has been obtained for REMVOW [propen-2-yl]-CH=C=Ru(PCy3)2Cl2.

The manual inspection of all the optimized geometries revealed incorrect PM7 coordination center geometries of some structures to be responsible for the PM7 failures. One particular example is conformer 5 of ruthenium complex REMVOW [propen-2-yl]-CH=C=Ru(PCy3)2Cl2. As depicted in Figure 7, PM7 geometry optimization of this structure resulted in Cl migration from Ru atom to carbon atom bonded to =C group of ruthenium took place manifesting chemical transformation. The subsequent PBE-D3/def2-svp optimization did not return Cl atom back to Ru to lead to initial structure. In contrast to conformer 5, all other conformers of REMVOW did not undergo to chemical transformation upon PM7 optimization. Similarly, the chemical transformation occurred upon PM7 geometry optimization of some (or all) conformers turned out to be responsible for large energy deviations obtained for other complexes, in particular for TITVEX01 (Ph₃P)₃AuGeCl₃ (GeCl₃ group dissociation from Au, see Figure S1), FODBOP [-CH=CMe-CH=CMe-]>Ir(H)(PEt₃)₃ (CH₂ dissociation from Ir for conformer 6, see Figure S2), LEHJAL [C₂H₅₃N₅O]₂PdCl₂ (coordination of the two hydrogen atoms to Pd, see Figure S3), MIYBAZ [N,N-diBu-benzimidazoline]₂PdI₂ (coordination the CH₂ groups to Pd, i.e. change square planar to octahedral configuration, see Figure S4), WOWSAC [MeOPh-Te-(CH₂)₂-Morpholinyl]₂PdCl₂ (hydrogen coordination to Pd, see Figure S5), CAFKOJ [(MeO)₃P]₂Ir(CO)
(OMe transfer from P to Ir for conformer 8, see Figure S6), EGOZUV (Furan)$_2$Se(CH$_2$SiMe$_2$Ph)$_3$
(dissociation of the two (conformer 2) or one (conformer 5) THF molecules from Sc, Figure S7),
CESHIS [C$_{14}$H$_{17}$NO$_4$]=C=Ru(PC$_3$)$_2$Cl$_2$ (Cl transfer from Ru to =C group for conformer 9, see
Figure S8). At the same time, correctly reproduced coordination center geometry does not
guarantee accurate relative conformational energies, as in case of ABAKOE (Ph-N-
iPr)$_3$W(=O)Cl.

An analysis of Figures 5 and 6 combined with manual geometries checks suggests that only for a
few (e.g. FONLID Hg[CH(COOEt)$_2$]$_2$, REPFID (HexylCyt)$_2$CdBr$_2$, LUWRUQ (Et$_2$N)$_3$V=O-
B(C$_6$F$_5$)$_3$) out of 27 structures PM7 method can be used to obtain the reliable conformational
energies and molecular geometries.

Figure 7. The molecular geometries of conformer 5 of REMVOW complex [propen-2-yl]-
CH=C=Ru(PC$_3$)$_2$Cl$_2$: initial PBE-D3 optimized conformer (left), the same conformer after PM7
gometry optimization (middle), the PM7 structure after subsequent PBE-D3 optimization
(right). Color coding: Ru (Orchid), Cl (Aquamarine), C (Turquoise), P (Orange). Hydrogens of high importance are transparent blue, otherwise omitted for clarity.

### 3.2.2 Geometry Optimization with Fixed Coordination Center

To see whether the constrained geometry optimization with all the atoms forming coordination center fixed at their DFT positions will improve on PM7 performance, we re-optimized all the PBE-D3 conformers obtained in Part I according to the procedure described in Section 2.2.2. The $\rho^2$ and $\rho$ values for correlation between PM7//PM7 and M06/def2-tzvp//PBE-D3/def2-svp relative conformational energies are presented in Figure 8. The average $\rho^2$ value obtained for all 27 structures turned out to be 0.46 which is practically identical to PM7 $\rho^2$ value of 0.45 obtained based on only SP energies (Part I). An average $\rho$ value turned out to be 0.37 which is reasonably close to PM7 $\rho$ value of 0.44 based on the SP energies and is higher than PM7 $\rho$ value of 0.25 obtained upon completely unrelaxed PM7 geometry optimizations. The largest $\rho^2$ values followed by positive $\rho$ values have been obtained for REPFID (HexylCyt)$_2$CdBr$_2$ ($\rho^2/\rho=0.77/0.88$), FONLID Hg[CH(COOEt)$_2$]$_2$ ($\rho^2/\rho=0.89/0.95$), FEGGII (-CH$_2$-)C(CH$_2$OBn)$_2$-(CH=CH$_2$)-Ir(PPh$_3$)$_2$(CO) ($\rho^2/\rho=0.81/0.90$), FODBOP [-CH=CMc-CH=CMc-]$_2$Ir(H)(PET$_3$)$_3$ ($\rho^2/\rho=0.83/0.91$), DUGVEH [(Me$_3$SiCH$_2$)$_2$Te]$_3$RhCl$_3$ ($\rho^2/\rho=0.85/0.92$) and CESHIS [C$_{14}$H$_{17}$NO$_4$]=C=Ru(PCy$_3$)$_2$Cl$_2$ ($\rho^2/\rho=0.92/0.961$) indicating correlation. No large $\rho^2$ values followed by negative $\rho$ values were obtained, indicating certain improvement upon fixed coordination center geometry optimizations.
Figure 8. The $\rho^2$ and $\rho$ correlation coefficients obtained between PM7 relative conformational energies on PM7 geometries with coordination center fixed during optimization and M06/def2-tzvp relative conformational energies on PBE-D3/def2-svp optimized geometries with PM7 optimized geometries used as starting point.
Figure 9. The mean unsigned deviation (MUD) values obtained between the PM7//PM7 (PM7 optimization with coordination center fixed) and M06/def2-tzvp//PBE-D3//def2-svp relative conformational energies for transition metal complexes studied in the present work. The solid bars indicate the average MUD value, and the ends of the solid lines at each bar give the lowest and the highest absolute deviation obtained for particular complexes.
In Figure 9 the MUDs obtained between PM7//PME7 (optimization with fixed coordination center) and M06/def2-tzvp//PBE-D3/def2-svp conformational energies are presented for every complex together with the largest and lowest absolute deviations in the relative conformational energies. An average MUD turned out to be 4.9 kcal/mol which is 1.6 kcal/mol lower comparing to what was obtained for fully relaxed PM7 geometry optimizations in Section 3.2.1. The lowest MUD of 0.8 kcal/mol has been obtained for complex FONLID Hg[CH(COOEt)₂]₂ and the largest MUD of 29.7 kcal/mol has been obtained for CODDIJ DET<(PPh₂)₂>PdCl₂.

Again the manual inspection of all the geometries has been performed. As all atoms forming the coordination center have been fixed during PM7 geometry optimization, distorted coordination center is not an origin of poorly predicted relative conformational energy. However, fixed coordination sphere does not eliminate another problem revealed for semiempirical methods in Section 3.2.1: incorrectly predicted coordination of some elements or functional groups to transition metal. Thus, during the PM7 geometry optimization of CODDIJ DET<(PPh₂)₂>PdCl₂ an unphysical coordination of one hydrogen atom in conformers 1, 2, 3, 5, 6, 8, 9 (see Figure S9 for conformer 1) or two hydrogen atoms in conformer 7 (see Figure S10) to Pd atom occurred. As no unphysical hydrogen coordination to Pd observed in conformers 0 and 4, these turned out to be significantly higher in energy according to PM7. DFT optimization removes false hydrogen coordination leading to significantly smaller energy diversity in conformational energies of CODDIJ DET<(PPh₂)₂>PdCl₂. Similar unphysical coordination of atoms/functional groups have also been found in CAFKOJ [(MeO)₃P]₃Ir(CO) (O of O-CH₃ group coordinates to Ir, see Figure S11), LEHJAL [C₂₁H₃₄N₅O₂]₄PdCl₂ (H of CH₃ groups coordination to Pd, see Figure S12), MIYBAZ [N,N-diBu-benzimidazoline]₂Pdl₂ (H of CH₃ coordination to Pd, see Figure S13), WOWSAC [MeOPh-Te-(CH₂)₂-Morpholinyl]₂PdCl₂ (H of CH₂ group coordination to Pd, see
Figure S14), REMVOW [propen-2-yl]-CH=Cr=PCy$_3$Cl$_2$ (H of CH$_2$ group coordination to Ru, see Figure S15). Remarkably, large errors have also been obtained for complexes for which no unphysical coordination to metal atoms occurred, e.g. for ABAKOE (Ph-N-iPr)$_3$W(=O)Cl, DEFVIT [(MeO)$_3$P]$_3$Cr(CO)$_2$(CSe), RESCUN [C$_{11}$H$_9$N$_2$]-CH$_2$-O-Zr(Furan)Cl$_4$ indicating fundamental difference between the PES in DFT and semiempirical methods\textsuperscript{49} which sometimes results in a conformational change, see Figures S16-S18.

4. Conclusion

A set of contemporary PM6*/7 black-box semiempirical methods has been tested to reproduce the conformational energies of 27 realistic transition metal complexes of 16 transition metals related to homogeneous catalysis. An analysis of the conformational energies derived from the single point energy evaluations revealed a pronounced difference between semiempirical and DFT methods. While conformational energies obtained with DFT protocols perfectly group together, their PM6*/7 counterparts turned out to be significantly different for all but a few complexes. To identify an origin of the deviations, we re-optimized all the conformers with PM7 method followed by DFT optimization mimicking the conformational search procedure in organometallic chemistry. Comparison of thus obtained PM7-based and DFT-based relative conformational energies revealed large discrepancies for all but a few complexes, originated from fundamental differences in PESs often manifesting themselves in chemical transformations/distortion of coordination center geometry or false coordination of some atoms (H, O) to transition metals upon PM7 geometry optimization. To minimize these effects, we tried
to perform PM7 geometry optimizations with fixed positions of all atoms composing the
coordination sphere. Even if slight improvement in PM7 performance was achieved, still large
discrepancies with DFT conformational energies have survived partly because fixing
coordination center did not prevent false coordination of atoms to transition metal to saturate the
coordination sphere. In cases where no false coordination was observed, an origin of the
difference in conformational energies is the fundamental difference between PESs of DFT and
semiempirical methods. Hence, in general the semiempirical methods are recommended for
conformational search/sampling in transition metal complexes only after careful validation for
particular complex. One possible origin of poor performance of the PM6*/PM7 semiempirical
methods is related to a limited number of complexes in the MOPAC training set,141 often
complexes containing necessary diatomic parameters are missing. For example, poor PM7
geometries obtained for TITVEX01 (Ph₃P)₃AuGeCl₃ complex might be the result of absence of
reliable Au – Ge diatomic parameters as there are no complexes both containing Au and Ge in
the PM7 training set.141 Another origin is likely related to the fact that many experimental
formation enthalpies tabulated in numerous databases as NIST webbook142-143 and utilized to
train semiempirical have been found to be inaccurate.144-146 We expect that careful re-
parameterization of semiempirical methods will significantly increase their accuracy for
transition metal systems. The obtained database147 is promising with respect to testing of other
strategies not covered by the present study as ligand field molecular mechanics (LFMM) method
developed by Deeth et al.67 and available via upcoming Tinker release68 where additional energy
terms describing the TM center are combined with standard force field terms responsible the
organic part of the molecule, and alternative quantum mechanically derived force field
(QMDF)148 or tight binding GNF-xTB36 approaches suggested by Grimme and co-workers.
ASSOCIATED CONTENT

Supporting Information. Cartesian coordinates (Å) of PBE-D3/def2-svp and PM7 optimized structures, M06/def2-tzvp, PBE0-D3/def2-tzvp, PBE-D3/def2-svp, PM6, PM6-D3, PM6-DH+, PM6-DH2, PM6-DH2X, PM6-D3H4, PM6-D3H4X and PM7, tabulated values forming the basis of Figures 2 – 6, Figures 8 – 9 and Figures S1-S18, tabulated conformational energies. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Yury.Minenkov@mipt.ru

*Luigi.Cavallo@kaust.edu.sa

ACKNOWLEDGMENT

We gratefully acknowledge Prof. Frank Jensen, Department of Chemistry, Aarhus University, Denmark, Prof. J. J. P. Stewart, Stewart Computational Chemistry, Colorado Springs, USA and Dr. Alexander M. Genaev, N. N. Vorozhtsov Institute of Organic Chemistry, Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia for helpful discussions. The research reported in this publication was supported by funding from King Abdullah University of Science and Technology (KAUST). 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. Y.M. gratefully acknowledges support from the Government of the Russian Federation (Agreement № 074-02-2018-286).
REFERENCES


87. Stewart, J. J. P., Optimization of Parameters for Semiempirical Methods. 3. Extension of PM3 to Be, Mg, Zn, Ga, Ge, As, Se, Cd, In, Sn, Sb, Te, Hg, Tl, Pb, and Bi. J. Comput. Chem. 1991, 12, 320-341.


For Table of Contents use only