The best density functional theory functional for the prediction of ^{1}H and ^{13}C chemical shifts of protonated alkylpyrroles

Zahn, Sarah L.V.; Hammerich, Ole; Hansen, Poul Erik; Sauer, Stephan P.A.

Published in:
Journal of Computational Chemistry

DOI:
10.1002/jcc.26540

Publication date:
2021

Document Version
Peer reviewed version

Citation for published version (APA):
Zahn, S. L. V., Hammerich, O., Hansen, P. E., & Sauer, S. P. A. (2021). The best density functional theory functional for the prediction of ^{1}H and ^{13}C chemical shifts of protonated alkylpyrroles. *Journal of Computational Chemistry*, *42*(18), 1248-1262. https://doi.org/10.1002/jcc.26540

The Best DFT Functional for the Prediction of ^{1}H and ^{13}C Chemical Shifts of Protonated Alkylpyrroles

Sarah L. V. Zahn [*], Ole Hammerich [†], Poul Erik Hansen [‡], Stephan P. A. Sauer [§]

April 8, 2021

Abstract

The prediction of ^{13}C chemical shifts can be challenging with density functional theory (DFT). In this study 39 different functionals and 3 different basis sets were tested on three neutral alkylpyrroles and their corresponding protonated species. The calculated shielding constants were compared to experimental data and results from previous calculations at the MP2. We find that the meta-hybrid functional TPSSh with either the Pople style basis set 6-311++G(2d,p) or the polarization consistent basis set pcSseg-1 gives the best results for the ^{13}C chemical shifts, whereas for the ^{1}H chemical shifts it is the TPSSh functional with either the 6-311++G(2d,p) or pcSseg-2 basis set. Including an explicit solvent molecule hydrogen bonded to NH in the alkylpyrroles improves the results slightly for the ^{13}C chemical shifts. On the other hand, for ^{1}H chemical shifts the opposite is true. Compared to calculations at the MP2 level none of the DFT functionals can compete with MP2 for the ^{13}C chemical shifts but for the ^{1}H chemical shifts the investigated DFT functionals are shown to give better agreement with experiment than MP2 calculations.

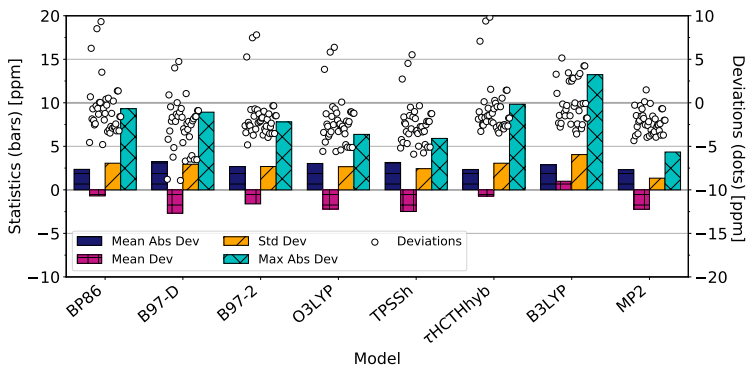
Keywords: DFT, Alkylpyrroles, Chemical Shifts, ^{13}C and ^{1}H NMR

[*]Department of Chemistry, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen, Denmark

[†]Department of Chemistry, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen, Denmark

[‡]Department of Science and Environment, Roskilde University, Universitetsvej 1, DK-4000 Roskilde, Denmark

[§]Department of Chemistry, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen, Denmark



Which DFT functional gives as good an agreement for ^{13}C and ^{1}H chemical shifts in neutral and protonated alkylpyrroles as MP2? 39 different functionals in combination with 3 different basis sets are systematically investigated. TPSSh in combination with the 6-311++G(2d,p) basis set is the winner. For the ^{1}H chemical shifts the agreement is even better than with MP2, while for the ^{13}C chemical shifts MP2 is still better.

INTRODUCTION

When determining the structure of an organic compound or distinguishing between different closely related structures, theoretical predictions of NMR spectra can be a useful tool for the organic chemist.[1–3] Many studies on calculating NMR chemical shifts, but almost none on cations, have been carried out to find a method which leads to good agreement with experimental values.[4,5] The methods and protocols suggested for calculations of chemical shifts, which is mostly density functional theory (DFT), often give large errors for cations of organic molecules. To successfully determine the chemical shifts for cations, second order Møller-Plesset perturbation theory (MP2) or Coupled Cluster theory are often the only methods that give results close to experimental data.[6–9] A recent study on protonated alkylpyrroles employing both DFT with the B3LYP functional and MP2 with various different solvent models concluded that B3LYP did not provide results in sufficient agreement with experimental results in contrast to MP2.[10] However, for larger compounds like e.g. oligomers of alkylpyrroles MP2 calculations are still computationally too demanding for routine applications, which makes it necessary to find cheaper alternatives. In the previous work, on the other hand, only the performance of the popular B3LYP exchange-correlation functional was investigated, which did not allow to make a general statement about the performance of density functional theory in such calculations. In the present work we discuss therefore the performance of in total 39 different exchange-correlation functionals in the calculation of both the ^{13}C and the ^{1}H chemical shifts of the same 6 alkylpyrrols. The aim of the study is to find a combination of exchange-correlation functional and basis set, which lead to results of sufficient accuracy. Such a method could then be used for calculations on mixtures of monomers and oligomers of protonated and neutral alkylpyrroles. The alkylpyrroles are of interest since they are important compounds for the synthesis of conducting polymeres and have been the focus of numerable studies.[11] With respect to the particular choice of molecules, it was important to demonstrate that the best solution found would cover both neutral and positively charged species and with the three compounds a broad range of substitution patterns was covered leaving an α-position for protonation.

COMPUTATIONAL DETAILS

The structures of the molecules investigated in this study are given in Chart 1 together with the acronyms by which they will be referred to throughout the paper.

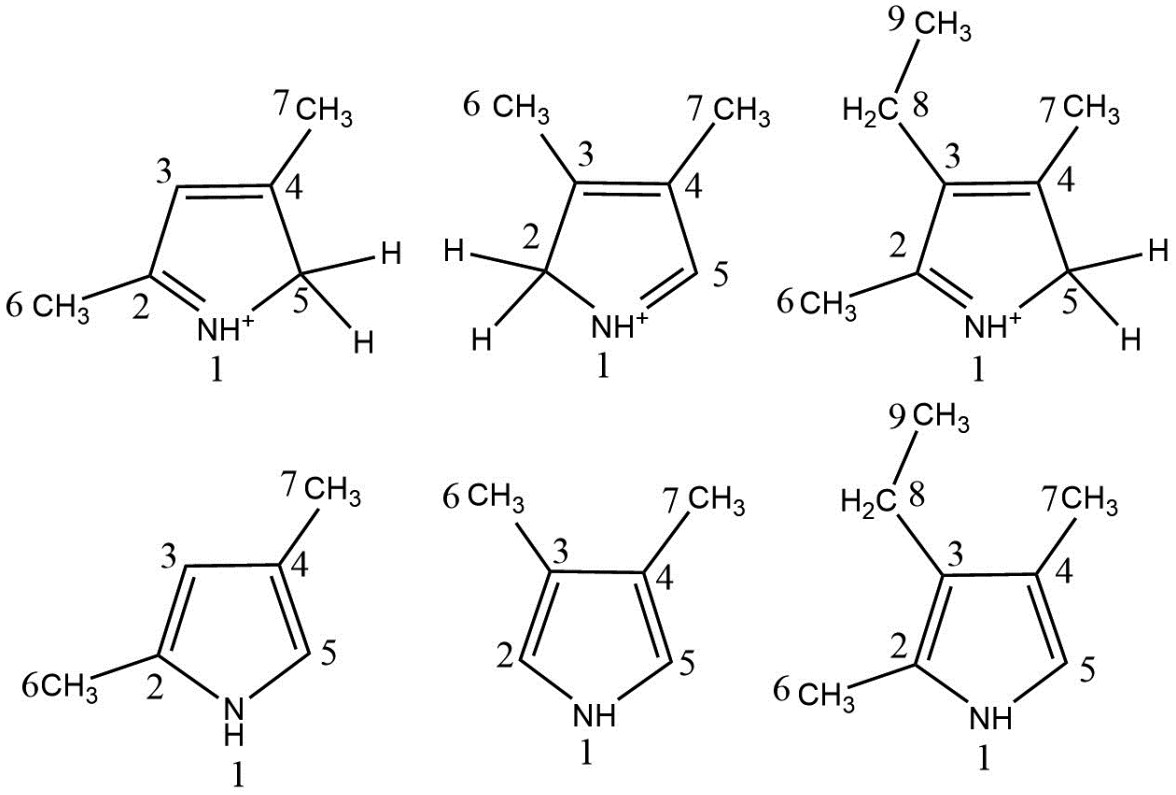


Chart 1: The six structures used in this study, from top left to bottom right 24dmp5H+, 34dmp2H+, 24dm3ep5H+, 24dmp, 34dmp and 24dm3ep.

39 different exchange-correlation functionals were tested in this study covering several of the steps of Jacob's ladder.[12] They include 9 generalized gradient approximation (GGA) functionals: BP86,[13,14] PBE,[15,16] B97-D,[17] KT1,[18] KT2,[18] KT3,[19] HCTH93,[20] HCTH147,[21] and HCTH407;[22] 16 hybrid GGA functionals: O3LYP,[23] B3P86,[13,14] B3PW91,[13,24] B3LYP,[13,25,26] mPW3PBE,[15,16,27] B97-1,[20] B97-2,[28] X3LYP,[29] PBE0,[30,31] mPW1PW91,[24,27] mPW1LYP,[26,27] mPW1PBE,[15,16,27] B98,[32,33] B1LYP,[34] PBEh1PBE[35] and BHandH;[25] 5 meta-GGA functionals: TPSS,[36] τHCTH,[37] VSXC,[38] M06-L,[39] M11-L;[40] 4 hybrid meta-GGA functionals: TPSSh,[36,41] τHCTHhyb,[37] M06-2X,[42] and B1B95[43] and 5 long-range corrected hybrid functionals: ωB97,[44] ωB97X,[44], ωB97X-D,[45] CAM-B3LYP[46] and LC-.ωPBE[47]

All geometry optimizations in this study were performed using the Gaussian 16 program package.[48] All shielding constants were calculated with the Gaussian 16 program package,[48] except for the ones done with KT1, KT2 and KT3, which were obtained using the Dalton 18 program package.[49,50] Three different basis sets were used for the geometry optimization: 6-311++G(2d,p),[51–53] pc-1 and pc-2.[54,55] For the NMR calculations 6-311++G(2d,p) was used, this basis set was also used for the geometry optimization, and pcSseg-1 and pcSseg-2[56] were employed, if pc-1 and pc-2 were used in the geometry optimizations, respectively. The polarizable consistent basis sets were all downloaded from the Basis Set Exchange database.[57,58] In the first study,[10] the cc-pVDZ basis set had been used for the geometry optimizations and the same geometry had been used in shielding calculations with different computational methods. In the present work, on the other hand, we used the more realistic and convenient scenario, i.e. that one uses the same functional and type of basis set for the geometry optimization as for the following shielding calculation. The start geometries for the current study were the optimized geometries of the conformer with the lowest energy from the previous study. This choice is also motivated by the fact the previous study had shown only a small influence of the details of the geometry optimization on the calculated shieldings.

All geometry optimizations and NMR calculations were performed using the Integral Equation Formalism Polarizable Continuum Model (IEFPCM)[59–61] with acetonitrile as solvent since the experimental data were obtained with acetonitrile as the solvent. The standard setting of Gaussian use UFF (Universal Force Field) for the atomic radii. The dielectric constants are Eps = 35.688 and Eps(inf) = 1.806874. The calculations were done for all six alkylpyrroles, and for each of the six compounds with and without an explicit solvent molecule hydrogen bounded to the NH moiety. The geometry optimizations were carried out with the "tight" convergence criteria (change in RMS density matrix $= 1.0 * 10^{-8}$, change in MAX density matrix $= 1.0 * 10^{-6}$, change in energy $= 1.0 * 10^{-6}$). The default Gaussian SCF algorithm DIIS was used as well as the default grid UltraFine with 99 radial shells and 590 angular points per shell. The Raffenetti 2 integral format was used with the Gaussian default accuracy of $1.0 * 10^{-12}$ for the 2-electron integrals. The absence of imaginary frequencies confirmed that they all led to true minima.

In the previous study[10] the comparison with experiment (Tables 1 - 6 in the supplementary material) was done by fitting the calculated shielding constants σ^{calc} to the experimental chemical shifts δ^{exp} according to equation (1):

$$\delta^{exp} = a\,\sigma^{calc} + b \tag{1}$$

and then judging the agreement based on the R^2 value of the fit and the slope a of the correlation. For the statistical comparison of the results from many functionals this is, however, not a convenient procedure. Therefore, in this study the data are treated differently. The atom with the most stable shielding constant considering all methods was choosen as internal reference, σ^{calc}_{ref} and δ^{exp}_{ref}. For all the other atoms i, it was then calculated how much the calculated relative chemical shift, $(\sigma^{calc}_{ref} - \sigma^{calc}_{i})$, deviates from the corresponding experimental relative chemical shift, $(\delta^{exp}_{i} - \delta^{exp}_{ref})$. The deviations between experimental and theoretical relative chemical shifts

$$\Delta_i = (\sigma^{calc}_{ref} - \sigma^{calc}_{i}) - (\delta^{exp}_{i} - \delta^{exp}_{ref}) \tag{2}$$

for all the atoms i were then used to calculated mean deviations (ME), mean absolute deviations (MAE), maximum absolute deviations (MAXERR) and the standard deviation (STD) of the Δ_i's from the mean deviation for a given method. For the ^{13}C chemical shifts C6 in 24dm3ep5H+ was thus selected as internal reference atom and for the ^{1}H it was C5-H in 24dmp.

In the previous work it could be seen that placing one explicit solvent molecule, acetonitrile, hydrogen bonded to NH in the alkylpyrroles, in addition to using only PCM, improved the results slightly,[10] so this approach is also explored here and will be denoted as the PCM+1S model.

For the most correct comparison with experiment one should of course also include vibrational corrections.[62–67] However, since we are aiming for a rather cheap approach to reproduce the experimental chemical shifts also for larger organic molecules, carrying out vibrational correction calculations is not really an option. Furthermore, we are dealing with molecules in solution and how to deal with solvent effects and vibrational corrections at the same time is still an open question.[68]

RESULTS AND DISCUSSION

^{13}C Chemical Shifts

Performance of the basis sets

Two types of basis sets were tested, the Pople style polarized triple zeta basis set 6-311++G(2d,p), and both the double and triple zeta polarizable consistent basis sets pcSseg-1 and pcSseg-2 in combination with pc-1 and pc-2 for the geometry optimization, respectively. In Figure 1, their performance is compared for one functional of each group. However, the meta and meta-hybrid functionals are treated as one group. The best or one of the best performing functionals of each group was chosen, i.e. the functional with the smallest standard deviation when comparing to experimental data. For the GGA functionals, this is B97-D, for the meta and meta-hybrid functionals it is TPSSh,which not surprisingly is a meta-hybrid GGA functional. From the hybrid functionals O3LYP was chosen and from the long-range corrected hybrid functionals ωB97.

When looking at the figure, it can be seen that 6-311++G(2d,p) and pcSseg-1//pc-1 have approximately the same accuracy, with 6-311++G(2d,p) often being a little better. 6-311++G(2d,p) and pcSseg-1//pc-1 employ the same type of functions for the hydrogen atoms and non-hydrogen atoms, but 6-311++G(2d,p), which is a triple zeta basis set, applies more functions to each type of atom than the double zeta basis sets pcSseg-1//pc-1. This might explain why 6-311++G(2d,p) performs better. This could lead to the expectation that the triple zeta basis sets pcSseg-2//pc-2, which employ even more functions for each atom would perform even better. This is however not the case. pcSseg-2//pc-2 by far gives the worst results when comparing to the experimental data. The same behaviour was previously already observed for the B3LYP functional.[10] The only functional for which pcSseg-2//pc-2 improves the results compared to pcSseg-1//pc-1 is the KT functional. But the improvement is not enough to give reasonable results compared to the other functionals. The pcSseg-2//pc-2 functional also improves the standard deviation and the maximum absolute deviation slightly for the meta functional TPSS, when one explicit solvent molecule is used, and it improves the results for the meta functional τHCTH. Otherwise, the pcSseg-2//pc-2 basis in

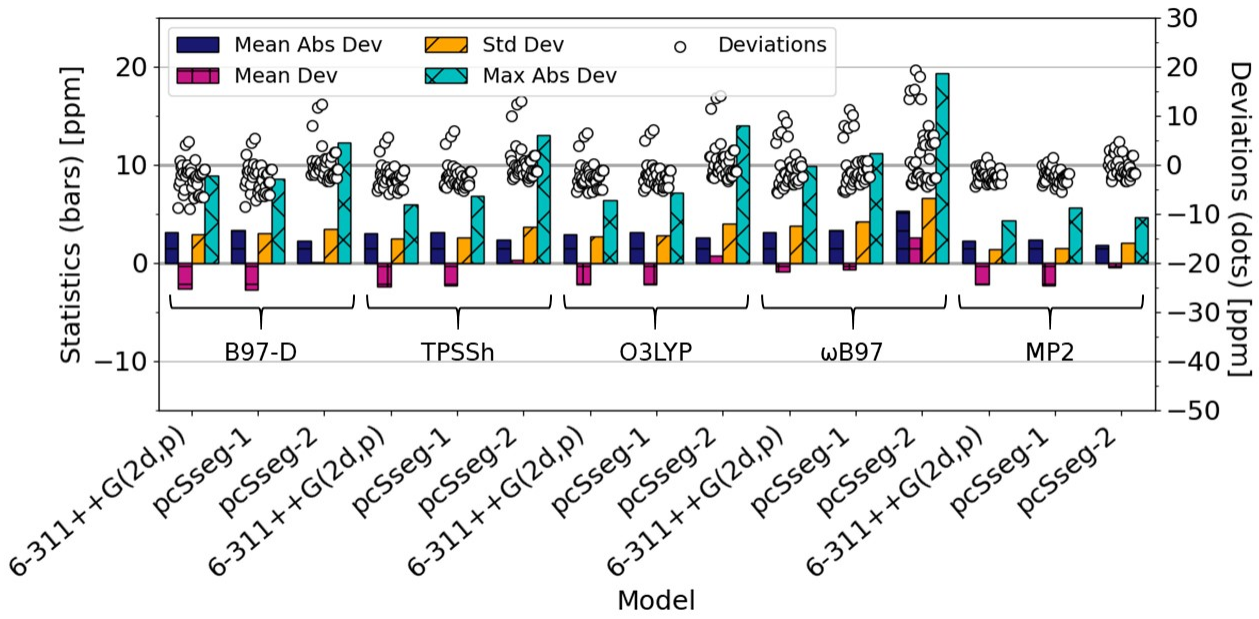


Figure 1: ^{13}C chemical shifts: performance of the different basis sets.

general overestimates the shielding constants considerably more than both it's double zeta version and the Pople style basis set. This trend can be seen also for all the other functionals. Figures and tables with the statistical values of all the functionals and basis set combinations can be found in the supplementary material (Tables 7 - 18 and Figures 1 - 12). The results are a bit different for MP2 where pcSseg-1//pc-1 for some parameters performs better than the Pople style basis set, and pcSseg-2//pc-2 performs better than pcSseg-1//pc-1 in the mean absolute deviation.

To conclude, either pcSseg-1//pc-1 or 6-311++G(2d,p) is recommended for the DFT calculations, and it is advised not to use pcSseg-2//pc-2. This conclusion is in good agreement with the previous article[10].

Performance of the functionals with PCM as solvent model and the 6-311++G(2d,p) basis set

First the best functional for each of the first categories of functionals will be determined when only PCM is used as solvent model. Two different approaches can be used when choosing the best functional from the statistical data. If the aim is to find the functional that performs best on average, then a mean deviation close to zero is important. This study, however, focuses on finding a method that will predict chemical shifts of all the individual atoms equally good. Therefore, it is more important that the standard deviation is low and that the mean absolute deviation and the absolute mean deviation are almost equal, because then an error correction can be applied to the results. For both approaches a low standard deviation and a low maximum deviation are both the aim, with standard deviation as the most important parameter. The individual deviations are also of great importance, since the functional with the fewest outliers – preferably one without outliers – can be considered to be the best performing functional.

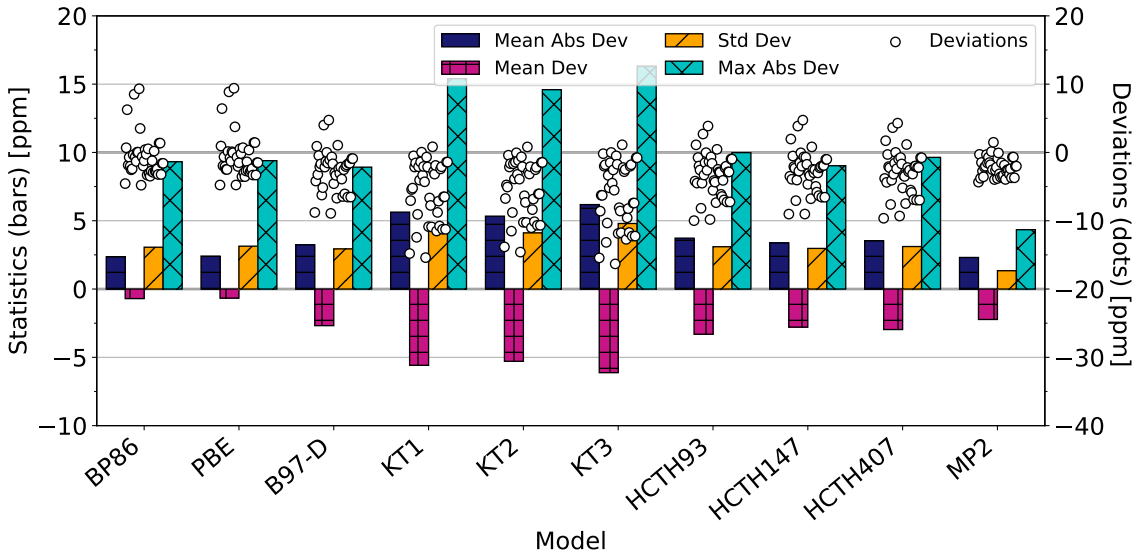


Figure 2: ^{13}C statistical data for the nine different GGA functionals and MP2, with the 6-311++G(2d,p) basis set. Experimental data was used as reference. Data for all six molecules is used for the statistical bars. Notice the difference in scaling on the two y-axes.

Figure 2 shows how well the different GGA functionals perform compared to each other and to MP2. The first thing to notice is that none of the GGA functionals is performing better than MP2, and that the Kohn-Tozer functionals, are the worst performing functionals when it comes to replicating the experimental data despite the fact, that they were actually optimized for properties like chemical shifts. The HCTH family of functionals gives very similar results, with HCTH147 performing slightly better, and HCTH93 the one from the family giving the results in worse agreement with experimental results. Looking at the deviations of chemical shifts of the individual carbon atoms (shown as open dots in the figure) BP86, PBE and B97-D are the best performing functionals. The deviations are more tightly grouped together for the BP86 and PBE functional with four noticeable outliers stemming from C4 in 24dm3ep5H+ with a deviation of 6.26 ppm for BP86 and 6.42 ppm for PBE, C4 in 24dmp5H+ with a deviation of 8.51 ppm for BP86 and 8.87 ppm for PBE, and C3 and C4 in 34dmp2H+ with deviations of 9.32 ppm and 3.51 ppm for BP86 respectively. For PBE the same atoms have deviations of 9.40 ppm and 3.75 ppm respectively. Please notice the different site of protonation in the latter molecule compared to the other two, which reverses the numbering of C3 and C4. The deviation of C3 is also the largest deviation for both BP86 and PBE. B97-D also has four noticeable outliers, but they are not stemming from the same four as for BP86 and PBE, C4 in 24dm3ep5H+ is no longer an outlier, it has a deviation of 0.91 ppm, whereas C2 has a deviation of -8.79 ppm, and is one of the negative outliers. C4 in 24dmp5H+ is still an outlier with a deviation of 4.01 ppm, and the same goes for C3 in 34dmp2H+ with a deviation of 4.74 ppm. The largest deviation, and the second negative outlier stems from C5 in 34dmp2H+, with a deviation of -8.92 ppm. In general there is no clear trend in which atoms are deviating the most for the GGA functionals. Comparing the statistical parameters of the three functionals and MP2 shows that MP2 has the best mean absolute deviation, standard deviation and maximum absolute deviation, with values of 2.31 ppm, 1.34 ppm and 4.34 ppm. The mean deviation for MP2 is -2.23 ppm which is numerically close to the mean absolute deviation, which is also good. The mean absolute deviation is lower for BP86 than for PBE and B97-D with values of 2.35 ppm, 2.41 ppm and 3.23 ppm, which makes sense when looking at the individual deviations. The mean deviation is also lower for BP86 and PBE, but the mean deviation is numerically closer to

the mean absolute deviation for B97-D with values of -0.70 ppm for BP86, -0.66 ppm for PBE and -2.68 ppm for B97-D. When it comes to standard deviation and maximum absolute deviation PBE now has the largest values of 3.14 ppm and 9.40 ppm, BP86 performs slightly better with values of 3.06 ppm and 9.32 ppm, whereas B97-D has values of 2.94 ppm and 8.92 ppm.

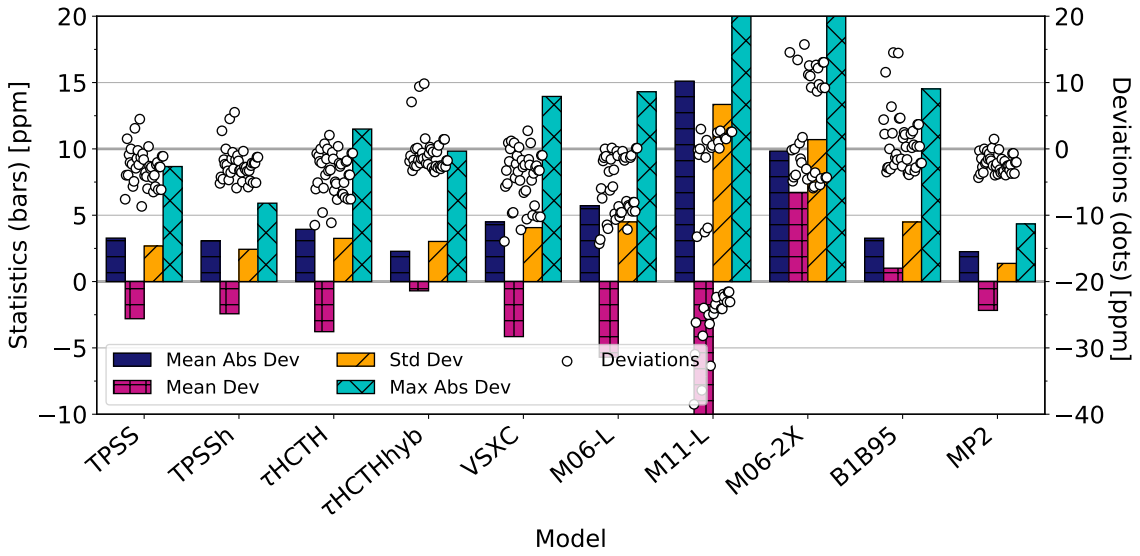


Figure 3: ^{13}C statistical data for the nine different meta/meta-hybrid GGA functionals and MP2, with the 6-311++G(2d,p) basis set. Experimental data was used as reference. Data for all six molecules is used for the statistical bars. TPSSh, τHCTHhyb, M06-2X and B1B95 is meta-hybrid GGA functionals. Mean deviation and maximum deviation for M11-L is -14.49 ppm and 38.50 ppm. Maximum deviation for M06-2X is 30.90 ppm. Notice the difference in scaling on the two y-axes.

When looking at Figure 3, which compares all the meta and meta-hybrid GGA functionals, it becomes very clear that the Minnesota functionals in general give some of the results in worst agreement with the experimental data. They overestimate the shielding constants. Overall the Minnesota family of functionals can therefore not be recommended for determining chemical shifts in alkylpyrroles. As for the GGA functionals, none of the functionals in this group is better than MP2. Looking at the figure there could also seem to be a trend that going from meta to meta-hybrid functional improves the result. When

going from TPSS to TPSSh the deviations become more tightly scattered, and the statistical parameters improve, the same goes for τHCTH and τHCTHhyb. Comparing the latter to the HCTH93, HCTH147 and HCTH407 functionals in Figure 2 also shows an improvement when going from GGA to meta-GGA. Looking at the individual deviations, both TPSSh and τHCTHhyb have three noticeable outliers which come from C4 in 24dm3ep5H+ and 24dmp5H+ with values of 2.73 ppm and 4.53 ppm for TPSSh and of 7.07 ppm and 9.39 ppm for τHCTHhyb. The third outlier stems from C3 in 34dmp2H+ with a value of 5.51 ppm for TPSSh and of 9.83 ppm for τHCTHhyb. The outliers have a much larger deviation for τHCTHhyb than for TPSSh, but the rest of the deviations are more tightly grouped together for τHCTHhyb. Unlike for the GGA functionals most of the meta/meta-hybrid functionals have the same three atoms as outliers. TPSSh, τHCTHhyb, B1B95 and even M06-2X (deviating to much to be seen on the figure) all have the same three atoms as outliers. TPPS have some negative outliers as well. Examining the statistical parameters shows that TPSSh has the lowest standard deviation and maximum absolute deviation with values of 2.43 ppm and 5.91 ppm compared to the values for τHCTHhyb of 3.06 ppm and 9.83 ppm. TPSSh has a larger mean absolute deviation, but a mean deviation numerically closer to the mean absolute deviation than τHCTHhyb has, with values for TPPSh of 3.36 ppm and -2.87 ppm compared to τHCTHhyb with values of 2.34 ppm and -0.72 ppm. The meta-GGA functional VSXC and the meta-hybrid-GGA functional B1B95 both have more scattered deviations than TPSS/TPSSh and τHCTH/τHCTHhyb, but they perform considerably better than M11-L and M06-2X. Their statistical parameters are also worse.

For the hybrid functionals in Figure 4 there is no pattern that shows any correlation between how much HF is used and how good the results are. BHandH which is the functional with most HF exchange included however is the worst performing hybrid functional. B3LYP which was the functional tested in the original study[10] is not one of the best performing functionals. Two functionals give better result than the rest, O3LYP and B97-2, but none of them are better than MP2. The deviations for O3LYP are more scattered than for B97-2, but the outliers are deviating more for B97-2 than for O3LYP. The outliers stem from C4 in 24dm3ep5H+ and 24dmp5H+ with values of 3.84 ppm and 5.82 ppm for O3LYP and 5.27 ppm and 7.47 ppm for B97-2. The third outlier for both functionals stems from C3

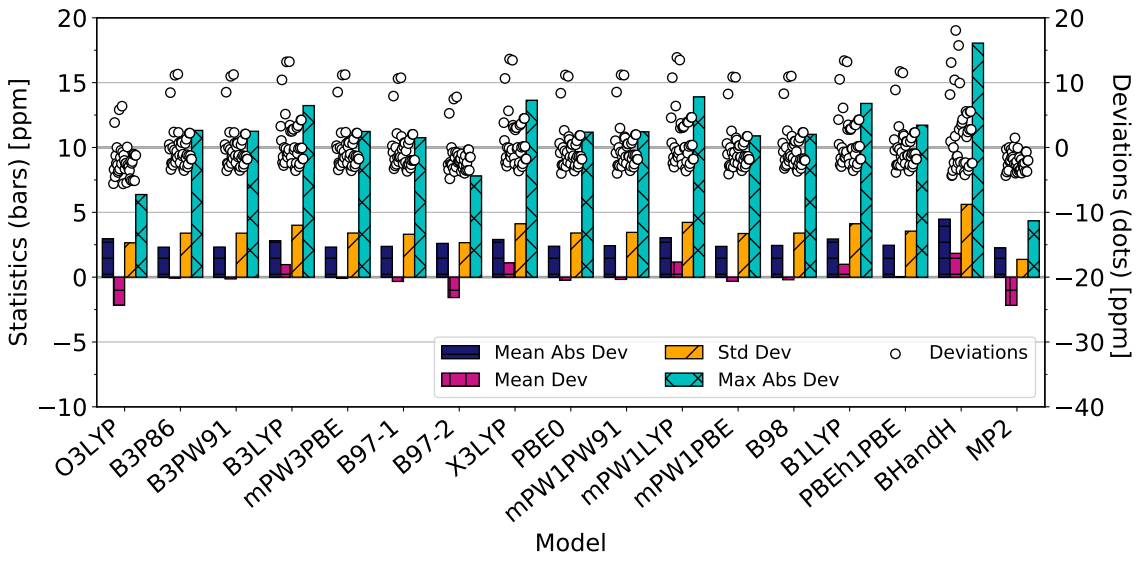


Figure 4: ^{13}C statistical data for the 16 different hybrid functionals, with the 6-311++G(2d,p) basis set. Experimental data was used as reference. Data for all six molecules is used for the statistical bars. The functionals was sorted after how much HF is included, from left O3LYP with 0.20 to right BHandH with 0.50 HF. Notice the difference in scaling on the two y-axis.

in 34dmp2H+ with a value of 6.37 ppm for O3LYP and of 4.74 ppm for B97-2. All other functionals have the same atoms as the three atoms with the largest deviations, including BHandH, that has some extra atoms with considerable deviations. Comparing the statistical parameters for the two functionals shows that their results are very close, with a mean absolute deviation of 3.03 ppm for O3LYP and 2.67 for B97-2, mean deviations of -2.20 ppm and -1.62 ppm, standard deviations of 2.66 ppm and 2.68 ppm and maximum absolute deviation of 6.37 for O3LYP and of 7.80 for B97-2. All other hybrid functionals have a standard deviation of 3 ppm or higher.

Figure 5 shows that all five long-range corrected functionals give almost the same accuracy, with the ωB97 family performing a little better than LC-ωPBE and the otherwise very often used functional CAM-B3LYP. But none of them gives better results than MP2. Comparing the long-range corrected hybrid functionals with the best performing functionals from the other categories shows that they have more distinctive outliers and all have larger

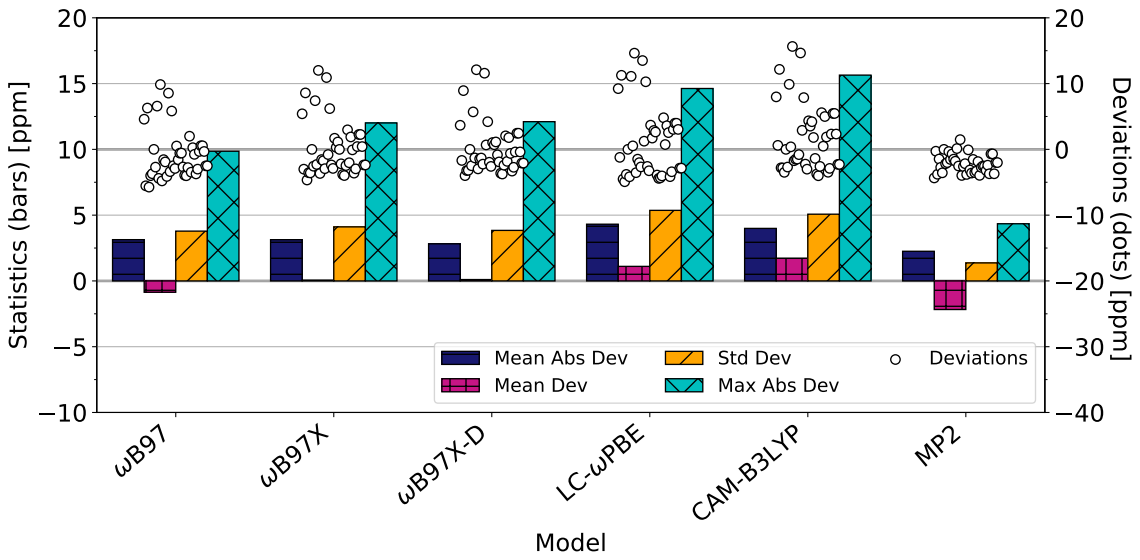


Figure 5: ^{13}C statistical data for the five different long-range corrected hybrid functionals and MP2, with the 6-311++G(2d,p) basis set. Experimental data was used as reference. Data for all six molecules is used for the statistical bars. Notice the difference in scaling on the two y-axis.

standard deviations and maximum absolute deviations than the best performing functionals in the other categories. They also all have larger mean absolute deviations than most of the best performing functionals from the other categories and they have mean errors numerically far from their mean absolute deviations. The six distinctive outliers for all five longe-range corrected hybrid functionals all stem from the same six atoms. How much each atom deviates changes a bit from functional to functional. The six atoms with the largest deviations are C2 and C4 in 24dm3ep5H+ and 24dmp5H+ and C3 and C5 in 34dmp2H+.

Statistical data for the six best performing functionals are collected in Figure 6 and Table 1. None of the long-range corrected functionals are among the six best performing functionals. Looking at Figure 6 it is important to notice that the three largest deviations for B3LYP are not shown, they come from C4 in 24dm3ep5H+ and 24dmp5H+ with values of 10.42 ppm and 13.21 ppm and from C3 in 34dmp2H+ with a value of 13.21 ppm, which is consistent with the findings of the previous study[10]. Many of the functionals have the same three atoms as outliers, some are able to give reasonable results for those atoms but

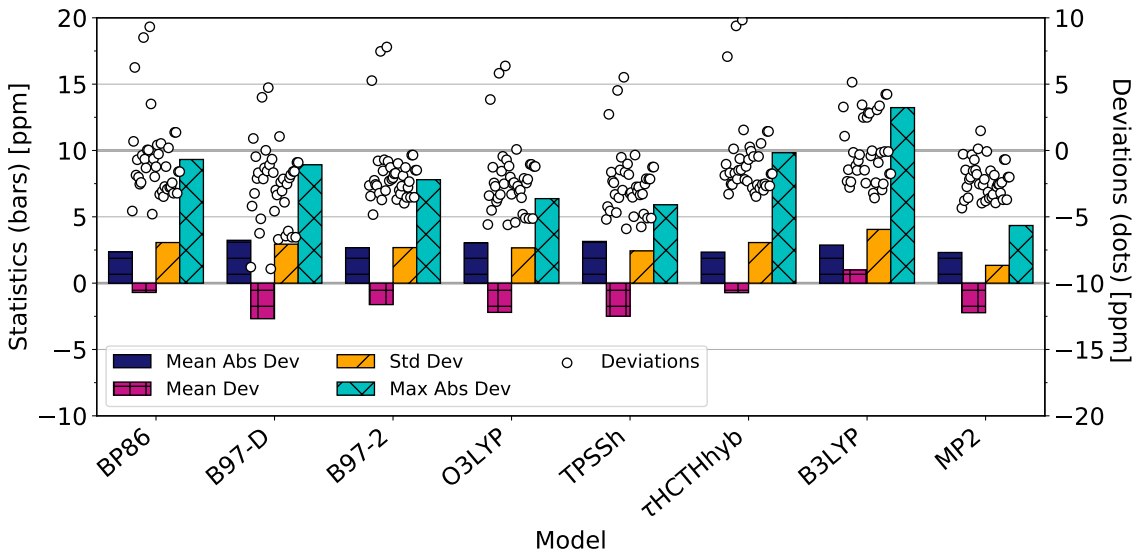


Figure 6: Statistical data for the six best functionals as well as B3LYP and MP2, with the 6-311++G(2d,p) basis set and PCM as solvent model. Experimental data was used as reference. Data for all six molecules is used for the statistical bars. Three largest deviations for B3LYP, which are not shown are 13.23 ppm, 13.21 ppm and 10.42 ppm

in turn have issues with other atoms such as C2 in 24dm3ep5H+ and 24dmp5H+. For 24dm3ep5H+ the chemical shift of C2 can both be overestimated and underestimated. C3 and C5 in 34dmp2H+ is also a challenge for some functionals. What they all have in common, is that the outliers always stem from the protonated molecules as observed previously.[10]. Furthermore the rest of the deviations for B3LYP are considerably more scattered than for especially B97-2, τHCTHhyb and MP2. Comparing the statistical data found in Table 1 for B3LYP with the six best functionals also shows that B3LYP is not the best choice. It has a larger standard deviation than all of them and a mean absolute deviation larger than BP86, B97-2 and τHCTHhyb. Furthermore, B3LYP has a larger numerical difference between mean absolute deviation and mean deviation than all of the best performing functionals and a maximum deviation much larger than the other functionals.

Table 1: ^{13}C statistical data for the six best performing functionals as well as B3LYP and MP2 when 6-311++G(2d,p) is used as basis set in combination with PCM as solvent model. Mean absolute deviation, mean deviation, standard deviation and maximum deviation are all given in ppm.

	BP86	B97-D	B97-2	O3LYP	TPSSh	τHCTHhyb	B3LYP	MP2
MAE	2.3551	3.2313	2.6704	3.0297	3.1457	2.3404	2.8716	2.3118
ME	-0.7022	-2.6815	-1.6172	-2.2034	-2.4908	-0.7172	0.9926	-2.2294
STD	3.0611	2.9376	2.6813	2.6589	2.4319	3.0600	4.0508	1.3412
MAXERR	9.3222	8.9233	7.8010	6.3716	5.9076	9.8314	13.2316	4.3452

Performance of the functionals with PCM and one explicit solvent molecule as solvent model and the 6-311++G(2d,p) basis set

When using the PCM+1S model, where one explicit solvent molecule hydrogen bonded to NH is used as an addition to the PCM model, the results improve slightly. This holds for all functionals. Figures corresponding to Figure 2-5 can be found in the supplementary material (Figures 13 - 24 and Tables 19 - 30). The conclusion on, which functional performs the best, is the same for the PCM+1S model as for the PCM model. The statistical data for the performance of the six best functionals and MP2 with the PCM+1S model can be found in Figure 7 and Table 2.

Comparing the results in Figure 7 with Figure 6 shows that for most of the functionals the deviations becomes less scattered, while for MP2 they become more scattered. Comparing the statistical parameters shows that the mean absolute deviation is improved for all functionals and for MP2. The numerical difference between mean absolute deviation and mean deviation also becomes smaller for all functionals except for BP86 where it is increased as it is also for MP2. The maximum absolute deviation is decreased for all functionals except for τHCTHhyb, the same goes for MP2. The improvement is however less than 1 ppm for all the statistical parameters except for the maximum absolute deviation, here BP86 is improved by 1.82 ppm, B97-2 by 1.43 ppm, τHCTHhyb by 1.63 ppm and B3LYP by 1.28 ppm. B97-2 and O3LYP both have improvements less than 0.30 ppm. Looking at the outliers for BP86,

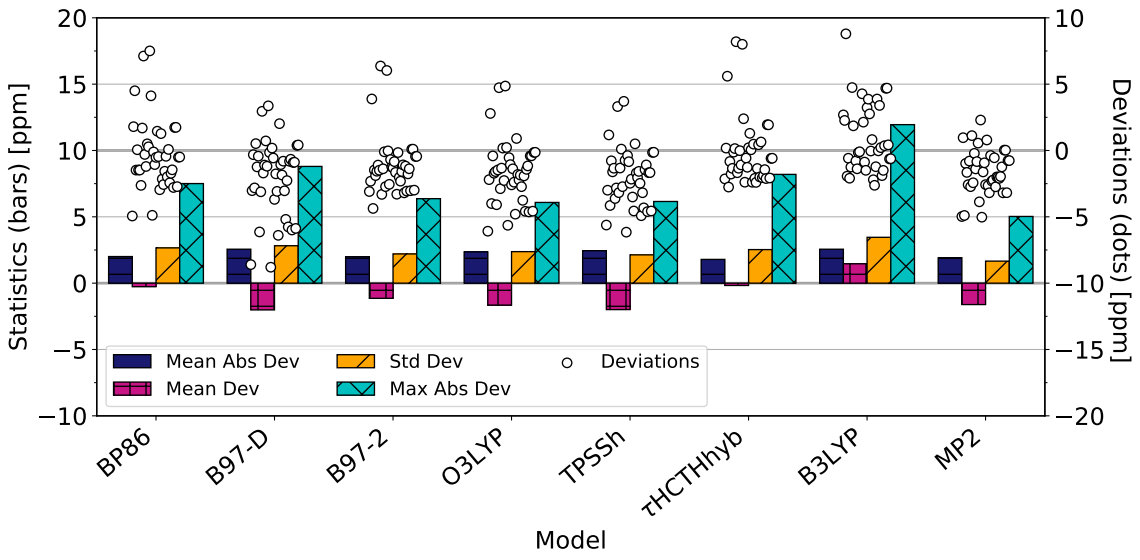


Figure 7: ^{13}C statistical data for the six best functionals as well as B3LYP and MP2, with the 6-311++G(2d,p) basis set and the PCM+1S solvent model. Experimental data was used as reference. Data for all six molecules is used for the statistical bars.

it can be seen that there are still four positive outliers stemming from the same four atoms. But now the values are 4.50 ppm for C4 in 24dm3ep5H+, 7.11 ppm for C4 in 24dmp5H+, 7.51 ppm for C3 and 4.12 ppm for C4 in 34dmp2H+, which means that the deviation for the first three outliers have been improved by more than 1 ppm, whereas the deviation for C4 in 34dmp2H+ has been increased by less than 1 ppm. The two negative outliers have not changed much, but the gap between them and the other deviations has increased. For B97-D there is no big change in the two negative outliers, but there are a few more atoms that could be considered as negative outliers. C4 in 24dmp5H+ now has a deviation of 2.95 ppm, and C3 in 34dmp2H+ has a deviation of 3.36 ppm and now C8 in 24dm3ep is also an outlier with a deviation of 2.02 ppm. The two original outliers have been improved by more than 1 ppm. For the hybrid functional B97-2 the three outliers still come from the same three atoms. However, it has changed, which of them that deviates the most. C4 in 24dm3ep5H+ now has a deviation of 3.88 ppm, C4 in 24dmp5H+ a deviation of 6.37 ppm and is now the atom with the largest deviation. C3 in 34dmp2H+ has a deviation of 6.03 ppm. For O3LYP it is also still C4 in 24dm3ep5H+ and 24dmp5H+ and C3 in 34dmp2H+ that are the three outliers; they now deviate by 2.79 ppm, 4.73 ppm and 4.86 ppm. For

TPSSh, C4 in 24dm3ep5H+ the deviation is now 1.17 ppm and is no longer an outlier. C4 in 24dmp5H+ and C3 in 34dmp2H+ are still considered outliers and have values of 3.31 ppm and 3.70 ppm. For τHCTHhyb, it are again the same three atoms that are considered outliers. C4 in 24dm3ep5H+ and 24dmp5H+ have deviations of 5.60 ppm and 8.20 ppm. C3 in 34dmp2H+ has a deviation of 8.00 ppm. For B3LYP the values are now 8.79 ppm, 11.95 ppm and 11.28 ppm. All original outliers for all functionals have been improved by more than 1 ppm.

Table 2: ^{13}C statistical data for the six best functionals as well as B3LYP and MP2 when 6-311++G(2d,p) is used as basis set in combination with the PCM+1S solvent model. Mean absolute deviation, mean deviation, standard deviation and maximum deviation are all given in ppm.

	BP86	B97-D	B97-2	O3LYP	TPSSh	τHCTHhyb	B3LYP	MP2
MAE	2.0050	2.5588	1.9886	2.3562	2.4445	1.7902	2.5579	1.9106
ME	-0.2689	-2.0175	-1.1438	-1.6541	-1.9936	-0.1751	1.4592	-1.6144
STD	2.6594	2.8199	2.2024	2.3760	2.1320	2.5299	3.4557	1.6604
MAXERR	7.5061	8.7991	6.3677	6.0840	6.1559	8.1980	11.9481	5.0277

Performance of the functionals with pcSseg-1//pc-1 as the basis set

When pc-1 was used for the geometry optimization and pcSseg-1 for calculating the shielding constants, the results did not change much. The six best functionals are still the same. The statistical data can be found in Figure 8 and Table 3. However, the deviations for most of the functionals are now scattered even more, especially the outliers have larger errors now. For B97-D, the other deviations are more grouped together than with the Pople style basis set, but the maximum deviations have increased. Looking at the statistical parameters they have all increased for all functionals, except for the mean absolute deviation for B97-D which is decreased by 0.90 ppm, and for τHCTHhyb, which is decreased by 0.03 ppm. The numerical difference between mean absolute deviation and mean deviation has decreased slightly for B3LYP while for all other functionals and for MP2 the numerical difference has increased.

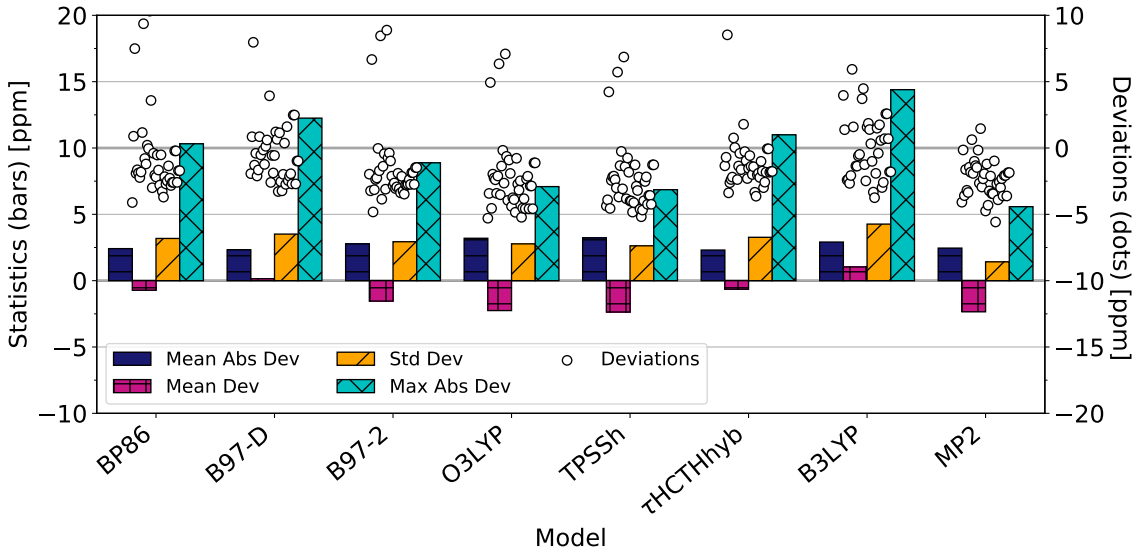


Figure 8: ^{13}C statistical data for the six best functionals as well as B3LYP and MP2, with the pcSseg-1//pc-1 basis set and PCM as the solvent model. Experimental data was used as reference. Data for all six molecules is used for the statistical bars. Not shown are the largest deviations for BP86 (10.32 ppm), the two largest deviations for B97-D (12.25 ppm and 11.70 ppm), the two largest deviations for τHCTHhyb (10.99 ppm and 10.38 ppm) and the rhree largest deviations for B3LYP (14.40 ppm, 14.18 ppm and 11.77 ppm).

Table 3: ^{13}C statistical data for the six best functionals as well as B3LYP and MP2 when pcSeg-1//pc-1 is used as basis set in combination with the PCM solvent model. Mean absolute deviation, mean deviation, standard deviation and maximum deviation are all given in ppm.

	BP86	B97-D	B97-2	O3LYP	TPSSh	τHCTHhyb	B3LYP	MP2
MAE	2.4153	2.3283	2.7697	3.1966	3.2235	2.3084	2.9111	2.4548
ME	-0.7218	0.1456	-1.5385	-2.2545	-2.3616	-0.6422	1.0335	-2.3478
STD	3.1828	3.5110	2.9342	2.7741	2.6279	3.2691	4.2641	1.4200
MAXERR	10.3181	12.2515	8.8848	7.0902	6.8585	10.9940	14.3968	5.5772

Adding one explicit solvent molecule does slightly improve the results for most of the functionals, but not enough to gain better results than adding one explicit solvent molecule with the Pople basis set as can be seen from the tables in the supplementary material (Tables 19 - 30 and Figures 13 - 24). For the GGA functionals, except for HCTH93 the standard deviation is improved slightly more when the pcSseg-1 basis set is used instead of the Pople basis set, the same goes for the meta functionals TPSS, VSXC and M06-L. Overall it is still better to use the Pople basis set than the pcSseg-1 basis set.

^{1}H Chemical Shifts

For hydrogen all atoms that could be considered to have the same chemical environment e.g. a methyl group have been treated as one. For a methyl group an average of the three computational values was calculated, and then used in the statistics only once.
The previous study showed that both MP2 and B3LYP had issues with replicating the chemical shifts for the NH protons and that the inclusion of an explicit solvent molecule improved the results for these atoms.[10] In the present study, the error from those atoms was, however, still very large and to get a more clear picture of the performance of the different functionals, it was chosen to omit these protons from the statistics. This approach can be justified by the fact that these type of protons are interchangeable with the solvent, and therefore their chemical shifts can also be hard to determine experimentally. Figures and tables of the result with the NH protons included can be found in the supplementary material (Figures 49 - 72 and Tables 55 - 78).

Performance of the basis sets

For ^{13}C there was a clear trend in the 6-311++G(2d,p) and pcSseg-1//pc-1 basis sets performing at approximately the same level of accuracy, and pcSseg-2 overestimating the chemical shifts far more. When examining Figure 9 where the results for the ^{1}H chemical shifts are shown, it can be seen that now it is the 6-311++G(2d,p) and pcSseg-2//pc-2 basis sets that are performing at approximately the same level of accuracy, with the Pople basis set performing a little better than pcSseg-2//pc-2. Figures and tables with the statistical values

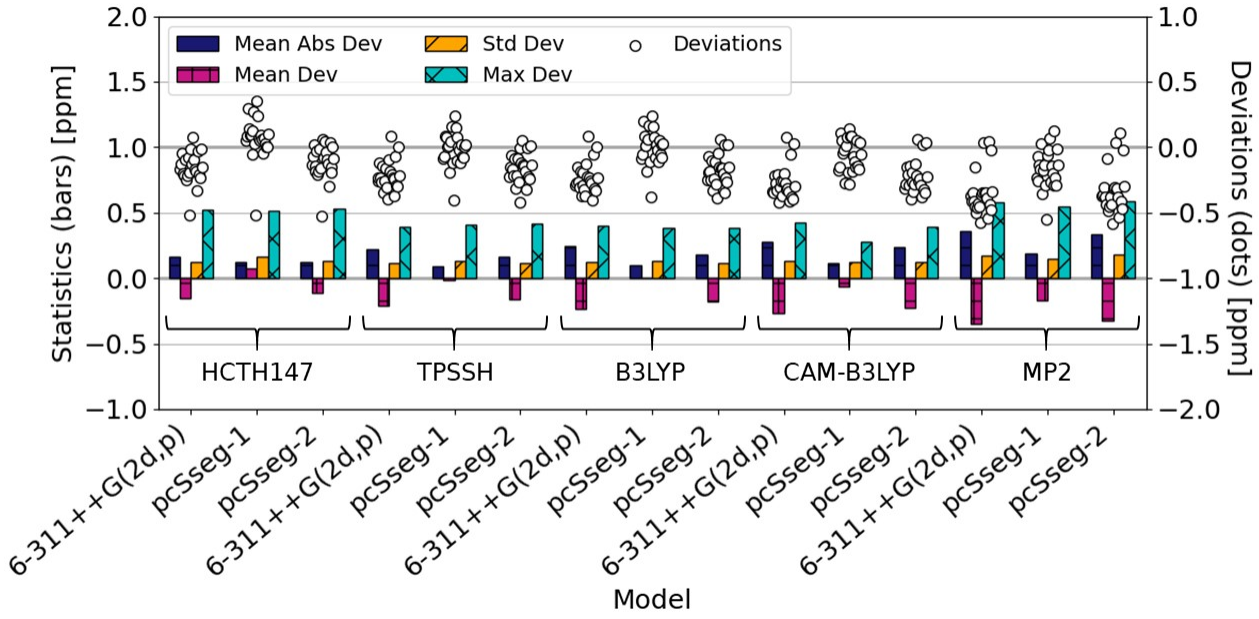


Figure 9: ^{1}H chemical shifts: performance of the different basis sets.

of all the functionals and basis set combinations can be found in the supplementary material (Tables 31 - 54 and Figures 25 - 48). This is thus not the same conclusion as for carbon, where it was Pople and pcSseg-1//pc-1. In the previous paper the basis set effects were already studied in greater detail employing also the even larger pcSseg-3 basis set. They found that the difference between the pcSseg-2 and pcSseg-3 was less than 1 ppm. The results however did not get closer to the experimental chemical shifts on going to the larger basis sets.[10] In short, convergence of the basis set does not necessarily guarantee results closer to experiment but will show the true performance of a functional. For smaller basis sets, one often can obtain then better agreement with experiment due to error cancellation but there is no guarantee that this will happen for the same basis sets for different atoms. In the following we will therefore only discuss results for the 6-311++G(2d,p) and pcSseg-2//pc-2 basis sets.

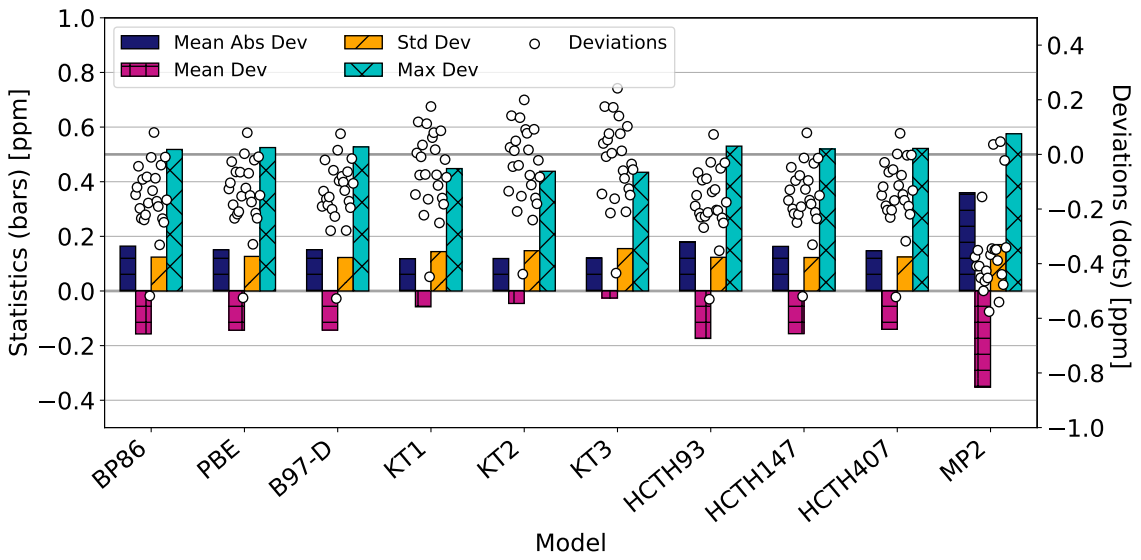


Figure 10: ^{1}H statistical data for the nine different GGA functionals and MP2, with the 6-311++G(2d,p) basis set. Experimental data was used as reference. Data for all six molecules is used for the statistical bars.

Performance of the functionals with PCM as solvent model and the 6-311++G(2d,p) basis set

The first thing to notice when going from ^{13}C to ^{1}H chemical shifts is that now MP2 is far from being the best method. In Figure 10 it can be seen that all the GGA functionals have lower standard deviations, mean absolute deviations and maximum absolute deviations than MP2. From the figure it can also be seen that the KT functionals have less trouble determining the ^{1}H chemical shifts than the ^{13}C chemical shifts. They all have some of the lowest maximum absolute deviations. However, they have their deviations more scattered than the other GGA functionals. The two functionals B97-D and HCTH147 both have the lowest standard deviation with a value of 0.120 ppm, but HCTH147 has a lower maximum absolute deviation with a value of 0.520 ppm compared to 0.528 ppm for B97-D. HCTH147 also has a smaller numerical difference between mean absolute deviation and mean deviation with a value of $6.9 * 10^{-3}$ ppm compared to B97-D with a value of $7.9 * 10^{-3}$ ppm. All the functionals have one distinctive outlier, stemming from the hydrogen on C5 in 34dmp2H+. This atom also gives rise to the maximum absolute deviation for all the GGA functionals. For

MP2. this atom is still the one that gives rise to the maximum absolute deviation. However, it is no longer considered an outlier. MP2 has three outliers compared to the deviations for the other hydrogen atoms with values close to 0 ppm stemming from hydrogen on C3 in 24dmp with a value of 0.048 ppm, from hydrogen on C5 in 24dm3ep with a value of 0.036 ppm and from hydrogen on C2/C5 in 34dmp with a value of -0.022 ppm. MP2 also have a fourth outlier stemming from C3-H in 24dmp5H+ with a value of -0.156 ppm. The best GGA functional for determining ^{1}H chemical shifts is HCTH147.

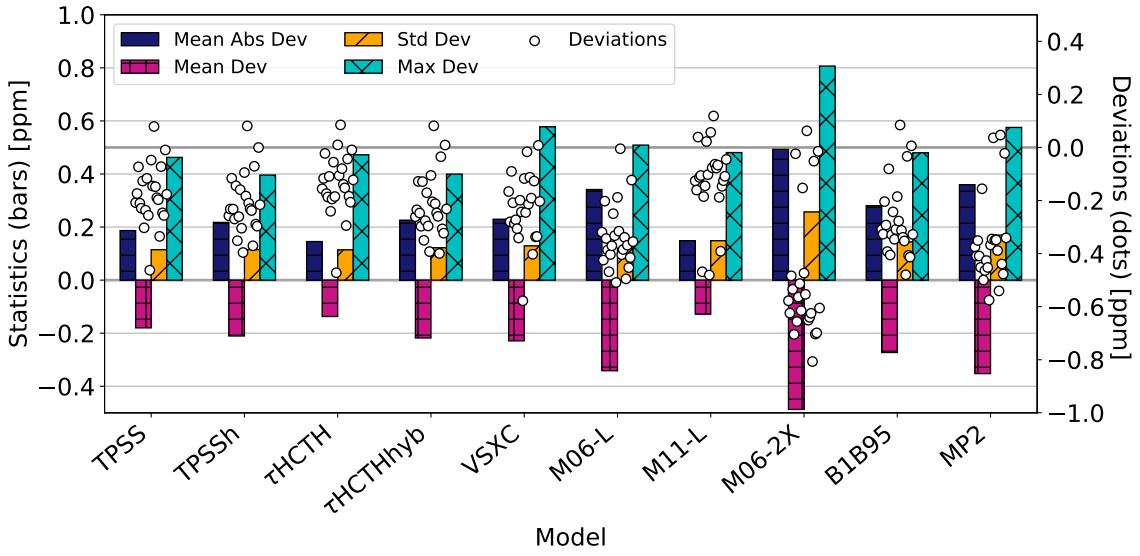


Figure 11: ^{1}H statistical data for the nine different meta/meta-hybrid GGA functionals and MP2, with the 6-311++G(2d,p) basis set. Experimental data was used as reference. Data for all six molecules is used for the statistical bars. TPSSh, τHCTHhyb, M06-2X and B1B95 is meta-hybrid GGA functionals.

Comparing Figure 3 with Figure 11 for the meta and meta-hybrid functionals another noticeable difference between carbon and hydrogen becomes clear. The Minnesota functionals were some of the worst performing functionals when determining ^{13}C chemical shifts, whereas M11-L is one of the meta functional with the lowest mean absolute deviation with a value of 0.148 ppm, when it comes to ^{1}H chemical shifts, and M06-L has no numerical difference between mean absolute deviation and mean deviation. M06-2X has still one of the largest standard deviation of 0.251 ppm, largest maximum absolute deviation with a value

of 0.807 ppm and the largest mean absolute deviation with a value of 0.493 ppm, but it has a small numerical difference between mean absolute deviation and mean deviation with a value of 0.0054 ppm. The functional with the lowest standard deviation is τHCTH with a value of 0.1116 ppm, TPSS and TPSSh comes very close with standard deviations of 0.112 ppm and 0.1126 ppm. τHCTH is also the meta-functional with the lowest mean absolute deviation, with a value of 0.145 ppm. TPSSh is the functional in the group with the lowest maximum absolute deviation with a value of 0.396 ppm, τHCTHhyb comes close with a value of 0.399 ppm. Looking at the individual deviations, TPSS has two that are considered outliers, one which is stemming from C5-H in 24dm3ep with a deviation of 0.079 ppm, and from C5-H in 34dmp2H+ with a deviation of -0.463 ppm, which is also the maximum absolute deviation. TPSSh however does not have C5-H in 34dmp2H+ as an outlier, but it still gives rise to the maximum absolute deviation. TPSSh does have C5-H in 24dm3ep as an outlier with a value of 0.081 ppm and C2/C5-H in 34dmp with a value of -0.0002 ppm. τHCTH has C5-H in 34dmp2H+ as the maximum absolute deviation and outlier with a value of -0.472 ppm. τHCTHhyb like TPSSh has C5-H in 24dm3ep as an outlier with a value of 0.082 ppm and C2/C5-H in 34dmp with a value of 0.0092 ppm, C6-H is also considered an outlier with a deviation of -0.105 ppm. C5-H in 34dmp2H+ is still the atom with the maximum absolute deviation. There could be a weak trend in going from meta to meta-hybrid functional in which type of atoms the functionals have trouble replicating the experimental data. The two other meta-hybrid functionals M06-2X and B1B95 also have C5-H in 24dm3ep and C2/C5-H in 34dmp as outliers. They do also have C3-H in 24dmp and 24dmp5H+ as an outlier just as MP2, and unlike most other functionals they have C9-H in 24dm3ep as their maximum absolute deviations. M06-2X has an extra outlier stemming from C5-H in 34dmp2H+. It is, howeve,r with a value 0f -0.153 ppm, which is very different from the other functionals, where this is an outlier. VSXC has C5-H in 34dmp2H+ as an outlier, which is also the maximum absolute deviation. M06-L has two distinctive outliers stemming from C5-H in 24dm3ep and C2/C5-H in 34dmp, and M11-L has 4 outliers close to 0 ppm stemming from 24dm3ep, C2/C5-H in 34dmp and C6-H in 34dmp2H+. M11-L also has 3 distinctive outliers stemming from C5-H in 34dmp2H+, which is also the maximum absolute deviation, and C3-H in 24dmp and 24dmp5H+. In general the accuracy of the meat and meta-hybrid func-

tionals varies much more than found for the GGA functionals. Some of them like M06-2X even gives results further away from the experimental values than MP2 does. The overall best functional from the group is TPSSh, when both statistical parameters and outliers are taking in consideration.

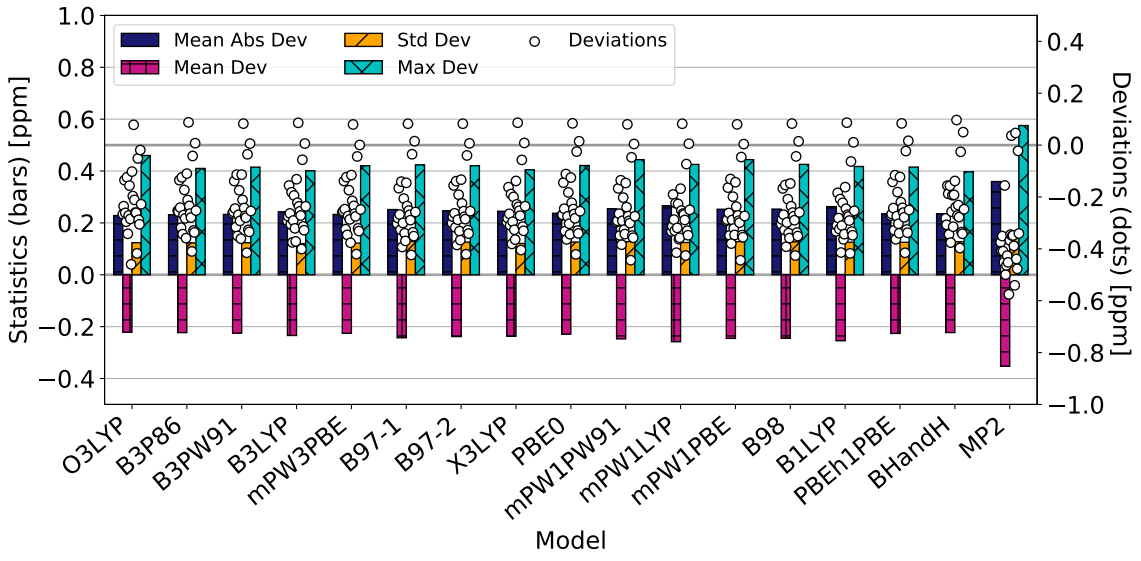


Figure 12: ^{1}H statistical data for the 16 different hybrid functionals, with the 6-311++G(2d,p) basis set. Experimental data was used as reference. Data for all six molecules is used for the statistical bars. The functionals was sorted after how much HF is included, from left O3LYP with 0.20 to right BHandH with 0.50 HF.

For the hybrid functionals, Figure 12, there is like for the carbon chemical sifts no correlation between the amount of HF exchange and how well the functionals perform. Unlike for carbon most of the hybrid functionals perform at the same level of accuracy, and all of them better than MP2. All the hybrid functionals have three distinctive outliers in form of C5-H in 24dm3ep, C3-H in 24dmp and C2/C5-H in 34dmp, they all have C9-H in 24dm3ep as their maximum absolute deviations, except for O3LYP which has C5-H in 34dmp2H+. The hybrid functional with the lowest standard deviation is B3LYP with a value of 0.1176 ppm. BHandH is the functional with the lowest maximum absolute deviation with a value of 0.3969 ppm and O3LYP is the functional with the lowest mean absolute deviation and the numerical difference between mean absolute deviation and mean deviation with values

of 0.2283 ppm and $6.8 * 10^{-3}$ ppm. B3LYP does have its deviations less scattered than O3LYP and has a lower maximum absolute deviation with a value of 0.4010 ppm, compared to 0.4599 ppm for O3LYP. X3LYP is the hybrid functional with the second lowest standard deviation with a value of 0.1180 ppm. The best hybrid functional is thus B3LYP.

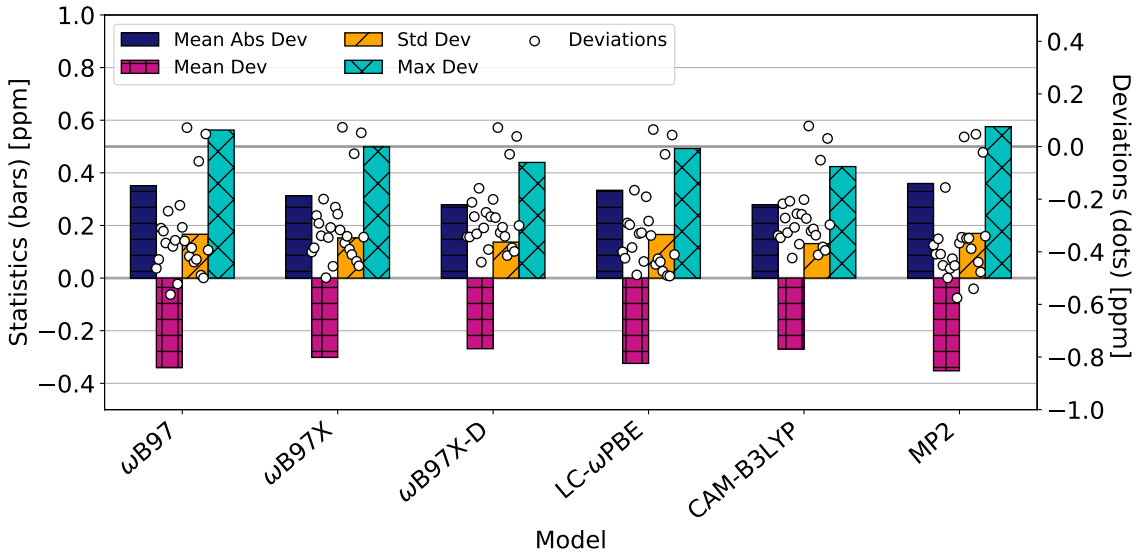


Figure 13: ^{1}H statistical data for the five different long-range corrected hybrid functionals and MP2, with the 6-311++G(2d,p) basis set. Experimental data was used as reference. Data for all six molecules is used for the statistical bars.

Figure 13 shows all the long-range corrected hybrid functionals and MP2. All of them perform better than MP2 on all statistical parameters, except for the numerical deviation between mean absolute deviation and mean deviation. CAM-B3LYP has the lowest standard deviation with a value of 0.1282 ppm, which is, however, larger than the standard deviation for all hybrid functionals. CAM-B3LYP is also the long-range corrected hybrid functional with the lowest mean absolute deviation and maximum absolute deviation with values of 0.2790 ppm and 0.4238 ppm. The long-range corrected hybrid functionals all have C5-H in 24dm3ep, C3-H in 24dmp and C2/C5-H in 34dmp as outliers and they all have C5-H in 24dm3ep5H+ as their maximum absolute deviation, except for LC-ωPBE which has C6-H as its maximum absolute deviation.

Figure 14 and Table 4 show the deviations and statistics for the seven best functionals and MP2. The reason for it being seven and not six as for the ^{13}C chemical shifts is due to

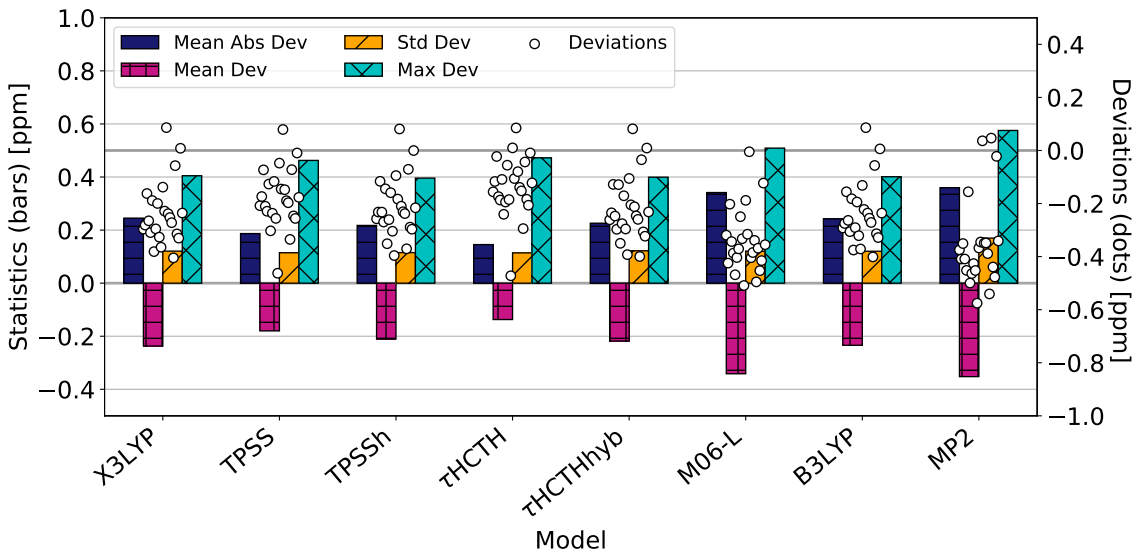


Figure 14: ^{1}H statistical data for the six best functionals, B3LYP and MP2, with the 6-311++G(2d,p) basis set. Experimental data was used as reference. Data for all six molecules is used for the statistical bars.

the fact that now B3LYP is one of the best functionals, and not just in the figure to compare with the results from the original paper. Looking at the figure the first thing to notice is that none of the GGA or long-range corrected functionals are part of this set of best performing functionals. Comparing the statistical data in Table 4 it can be seen they all have almost the same standard deviation as well as the mean absolute deviations and mean deviations close to each other, which are the two most important statistical parameters. Comparing how much the deviations are scattered, however, shows that TPSS and τHCTHhyb have some of the least scattered deviations, and since TPSSh has a slightly lower standard deviation and maximum absolute deviation, it is considered to be the best performing functional when it comes to replicating the ^{1}H chemical shift using the 6-311++G(2d,p) basis set and the PCM solvent model.

Table 4: ^{1}H: the six best functionals, B3LYP and MP2 when 6-311++G(2d,p) is used as basis set. Mean absolute deviation, mean deviation, standard deviation and maximum deviation are all given in ppm.

	X3LYP	TPSS	TPSSh	τHCTH	τHCTHhyb	M06-L	B3LYP	MP2
MAE	0.2451	0.1868	0.2176	0.1454	0.2259	0.3415	0.2422	0.3596
ME	-0.2369	-0.1799	-0.2105	-0.1371	-0.2180	-0.3415	-0.2342	-0.3523
STD	0.1180	0.1120	0.1126	0.1116	0.1193	0.1179	0.1176	0.1661
MAXERR	0.4047	0.4626	0.3959	0.4722	0.3994	0.5089	0.4010	0.5755

Performance of the functionals with PCM and one explicit solvent molecule as solvent model and the 6-311++G(2d,p) basis set

Unlike for ^{13}C adding one explicit solvent molecule hydrogen bonded the NH group makes the results worse compared to experimental data for the ^{1}H chemical shifts with the exception of the NH proton, which is not considered here. For all the GGA functionals the standard deviation and maximum absolute deviation is slightly increased and for all of them with the exception of the KT functionals the mean absolute deviation, however, is decreased. For the meta and meta-hybrid functionals, the mean absolute deviation is also decreased, except for the mean absolute deviation for M11-L. For M06-L, M11-L, M06-2X and B1B95 the standard deviation, and maximum absolute deviation are improved. For the hybrid functionals, the mean absolute deviation has also been decreased for all of them, but the standard deviation is increased. For most of them the maximum absolute deviation has also been increased, except for the B3P86, B3PW91, mPW3PBE, mPW1PW91 and mPW1PBE functionals, where it has been slightly decreased. The trend is also seen for the long-range corrected hybrid functionals, the mean absolute deviation has been slightly improved, and standard deviation and maximum absolute deviation have been slightly increased, except for the standard deviation for LC-ωPBE, which has been improved slightly. Of the few functionals whose performance has been improved by adding an explicit solvent molecule only M06-L has been improved enough to be of any importance. M06-L is now the functional with the lowest standard deviation and maximum absolute deviation with a value of 0.1025 ppm and

0.2401 ppm. The numerical difference between mean absolute deviation and mean deviation has however been increased from 0 ppm to 0.0664 ppm; the two parameters have though both been improved.

The results are another story when the NH protons are not omitted, here the results are much improved when an explicit solvent molecule is added. Figures and tables for this can be found in the supplementary material (Figures 49 - 72 and Tables 55 - 78).

Performance of the functionals with pcSseg-2//pc-2 as the basis set

With the pcSseg-2//pc-2 basis set the conclusion, as to which functionals are the best, changes a little, the functionals TPPSSh, X3LYP and B3LYP are still part of the best functionals. The statistical parameters and deviation for the seven best functionals and MP2 with the pcSseg-2//pc-2 basis set can be found in Figure 15 and Table 5. Figure 15 and Table 5 also show that MP2 is still the worst performing method when the basis set is changed to pcSseg-2//pc-2. Comparing the numbers for X3LYP in Tables 4 with the numbers in 5, it can be seen that unlike for most of the other functionals, the results actually improve when changing to the pcSseg-2//pc-2 basis set. The standard deviation for X3LYP is, however, still higher than for TPSSh both with the Pople basis set and with the pcSseg-2//pc-2 basis set. The maximum absolute deviation is however lower for X3LYP than for TPSSh with both basis sets. The mean absolute deviation is also still lower for TPSSh with both basis sets than it is for X3LYP. Most of the functionals in Figure 15 have C5-H in 34dmp2H+ as their maximum absolute deviation. For mPW1LYP and M06-L it is C9-H in 24dm3ep and for B1LYP it is C5-H in 24dmp5H+.

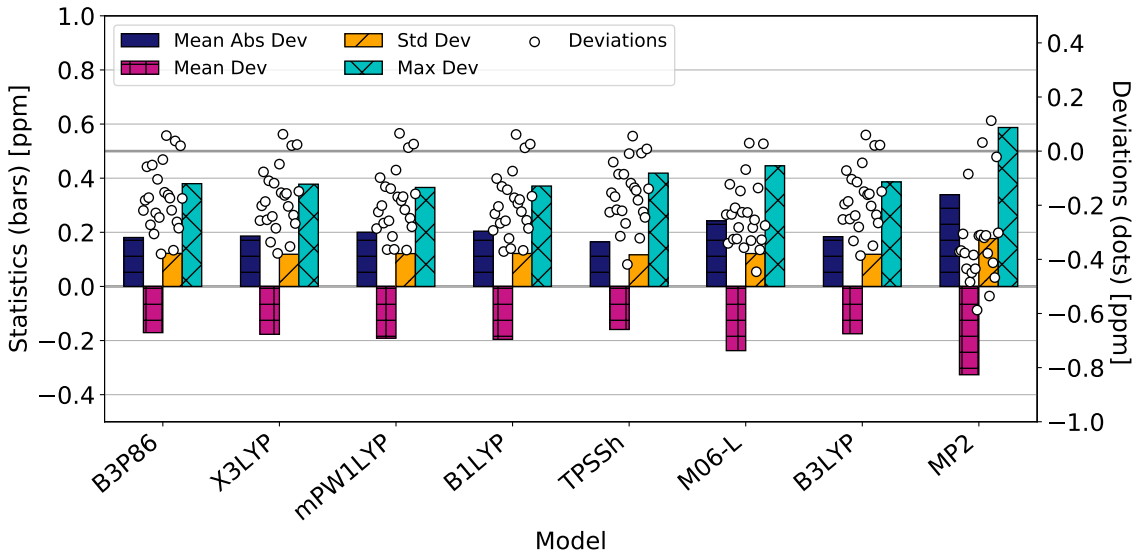


Figure 15: ^{1}H statistical data for the six best functionals, B3LYP and MP2, with the pcSseg-2 basis set. Experimental data was used as reference. Data for all six molecules is used for the statistical bars.

Table 5: ^{1}H: the six best functionals as well as B3LYP and MP2 when pcSseg-2//pc-2 is used as basis set. Mean absolute deviation, mean deviation, standard deviation and maximum deviation are all given in ppm.

	B3P86	X3LYP	mPW1LYP	B1LYP	TPSSh	M06-L	B3LYP	MP2
MAE	0.1808	0.1862	0.2004	0.2042	0.1647	0.2422	0.1839	0.3389
ME	-0.1708	-0.1769	-0.1913	-0.1956	-0.1591	-0.2373	-0.1749	-0.3263
STD	0.1195	0.1163	0.1187	0.1191	0.1143	0.1189	0.1162	0.1730
MAXERR	0.3797	0.3779	0.3657	0.3710	0.4187	0.4458	0.3862	0.5873

CONCLUSIONS

The aim of this study was to find a DFT method that was better than MP2 and B3LYP at replicating chemical shields for neutral and protonated alkylpyrroles. For the ^{13}C chemical shifts it was only possible to find a functional better than B3LYP, but not one better than MP2. Most of the functionals performed better than B3LYP. TPSSh is the best method

for determining ^{13}C chemical shifts either with the Pople style basis set 6-311++G(2d,p) or the polarization consistent basis set pcSseg-1//pc-1. The results are slightly better with the Pople style basis set. The results can be slightly improved by an explicit solvent molecule hydrogen bonded to the NH-moiety. Almost all the functionals struggle to accurately predict the chemical shifts of the carbon atoms carrying the positive charge in the protonated versions.

For the ^{1}H chemical shifts all the functionals performed better than MP2 and B3LYP was one of the best performing functionals. For hydrogen the best functional was likewise TPSSh with the 6-311++G(2d,p) basis set, but here the pcSseg-2//pc-2 basis set performed better than pcSseg-1//pc-1. If the chemical shift of both ^{1}H and ^{13}C is of interest the best method is TPSSh/6-311++G(2d,p) For the ^{1}H chemical shifts the results get slightly worse when adding an explicit solvent molecule. If the chemical shifts of the NH protons are needed, it does, however, improve the results significantly, since the results for those atoms get more accurate. The results when using an explicit solvent molecule do, however, not improve to the same degree as omitting the NH protons.

ACKNOWLEDGEMENTS

We thank Dr. Fadhil S. Kamounah for his help in producing a sample of the protonated 3,4dmpH+. SLVZ and SPAS thank the Department of Chemistry, University of Copenhagen for access to its high-performance computer cluster.

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. C. A. Hunter, M. J. Packer, and C. Zonta, Prog. NMR Spectrosc. **47**, 27 (2005).

2. M. W. Lodewyk, M. R. Siebert, and D. J. Tantillo, Chem. Rev. **112**, 1839 (2012).

3. L. B. Krivdin, Prog. NMR Spectrosc. **112-113**, 103 (2019).

4. K. Gholivand, Y. Maghsoud, M. Hosseini, and M. Kahnouji, J. Mol. Struc. **1183**, 230 (2019).

5. E. Benassi, J. Comp. Chem. **38**, 87 (2017).

6. H.-U. Siehl, T. Müller, and J. Gauss, J. Phys. Org. Chem. **16**, 577 (2003).

7. H.-U. Siehl, T. Mueller, J. Gauss, P. Buzek, and P. v. R. Schleyer, J. Am. Chem. Soc. **116**, 6384 (1994).

8. J. F. Stanton, J. Gauss, and H.-U. Siehl, Chem. Phys. Lett. **262**, 183 (1996).

9. M. E. Harding, J. Gauss, and P. v. R. Schleyer, J. Phys. Chem. A **115**, 2340 (2011).

10. E. G. Lacerda Jr, F. S. Kamounah, K. Coutinho, S. P. A. Sauer, P. E. Hansen, and O. Hammerich, ChemPhysChem **20**, 78 (2019).

11. S. Sadki, P. Schottland, N. Brodie, and G. Sabouraud, Chem. Soc. Rev. **29**, 283 (2000).

12. J. P. Perdew and K. Schmidt, AIP Conference Proceedings **577**, 1 (2001).

13. A. D. Becke, Phys. Rev. A **38**, 3098 (1988).

14. J. P. Perdew, Phys. Rev. B **33**, 8822 (1986).

15. J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).

16. J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **78**, 1396 (1997).

17. S. Grimme, J. Comp. Chem. **27**, 1787 (2006).

18. T. W. Keal and D. J. Tozer, J. Chem. Phys. **119**, 3015 (2003).

19. T. W. Keal and D. J. Tozer, J. Chem. Phys. **123**, 121103 (2005).

20. F. A. Hamprecht, A. J. Cohen, D. J. Tozer, and N. C. Handy, J. Chem. Phys. **109**, 6264 (1998).

21. A. D. Boese, N. L. Doltsinis, N. C. Handy, and M. Sprik, J. Chem. Phys. **112**, 1670 (2000).

22. A. D. Boese and N. C. Handy, J. Chem. Phys. **114**, 5497 (2001).

23. A. J. Cohen and N. C. Handy, Mol. Phys. **99**, 607 (2001).

24. J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, and C. Fiolhais, Phys. Rev. B **46**, 6671 (1992).

25. B. A. D, J. Chem. Phys. **98**, 5648 (1993).

26. C. Lee, W. Yang, and R. G. Parr, Phys. Rev. B **37**, 785 (1988).

27. C. Adamo and V. Barone, J. Chem. Phys. **108**, 664 (1998).

28. P. J. Wilson, T. J. Bradley, and D. J. Tozer, J. Chem. Phys. **115**, 9233 (2001).

29. X. Xu and W. A. Goddard, Proc. Natl. Acad. Sci. **101**, 2673 (2004).

30. M. Ernzerhof and G. E. Scuseria, J. Chem. Phys. **110**, 5029 (1999).

31. C. Adamo and V. Barone, J. Chem. Phys. **110**, 6158 (1999).

32. A. D. Becke, J. Chem. Phys. **107**, 8554 (1997).

33. H. L. Schmider and A. D. Becke, J. Chem. Phys. **108**, 9624 (1998).

34. C. Adamo and V. Barone, Chem. Phys. Lett. **274**, 242 (1997).

35. M. Ernzerhof and J. P. Perdew, J. Chem. Phys. **109**, 3313 (1998).

36. J. Tao, J. P. Perdew, V. N. Staroverov, and G. E. Scuseria, Phys. Rev. Lett. **91**, 146401 (2003).

37. A. D. Boese and N. C. Handy, J. Chem. Phys. **116**, 9559 (2002).

38. T. Van Voorhis and G. E. Scuseria, J. Chem. Phys. **109**, 400 (1998).

39. Y. Zhao and D. G. Truhlar, J. Chem. Phys. **125**, 194101 (2006).

40. R. Peverati and D. G. Truhlar, J. Phys. Chem. Lett. **3**, 117 (2012).

41. V. N. Staroverov, G. E. Scuseria, J. Tao, and J. P. Perdew, J. Chem. Phys. **119**, 12129 (2003).

42. Y. Zhao and D. G. Truhlar, Theor. Chem. Acc. **120**, 215 (2008).

43. A. D. Becke, J. Chem. Phys. **104**, 1040 (1996).

44. J.-D. Chai and M. Head-Gordon, J. Chem. Phys. **128**, 084106 (2008).

45. J.-D. Chai and M. Head-Gordon, Phys. Chem. Chem. Phys. **10**, 6615 (2008).

46. T. Yanai, D. P. Tew, and N. C. Handy, Chem. Phys. Lett. **393**, 51 (2004).

47. O. A. Vydrov and G. E. Scuseria, J. Chem. Phys. **125**, 234109 (2006).

48. 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, et al., *Gaussian 16 Revision 16.A.03* (2016).

49. K. Aidas, C. Angeli, K. L. Bak, V. Bakken, L. Boman, O. Christiansen, R. Cimiraglia, S. Coriani, P. Dahle, E. K. Dalskov, et al., WIREs Comput. Mol. Sci. **4**, 269 (2014).

50. J. M. H. Olsen, S. Reine, O. Vahtras, E. Kjellgren, P. Reinholdt, K. O. Hjorth Dundas, X. Li, J. Cukras, M. Ringholm, E. D. Hedegård, et al., J. Chem. Phys. **152**, 214115 (2020).

51. R. Krishnan, J. S. Binkley, R. Seeger, and J. A. Pople, J. Chem. Phys. **72**, 650 (1980).

52. M. J. Frisch, J. A. Pople, and J. S. Binkley, J. Chem. Phys. **80**, 3265 (1984).

53. M. M. Francl, W. J. Pietro, W. J. Hehre, J. S. Binkley, M. S. Gordon, D. J. DeFrees, and J. A. Pople, J. Chem. Phys. **77**, 3654 (1982).

54. F. Jensen, J. Chem. Phys. **115**, 9113 (2001).

55. F. Jensen, J. Chem. Phys. **116**, 3502 (2002).

56. F. Jensen, J. Chem. Theory Comput. **4**, 719 (2008).

57. D. Feller, J. Comp. Chem. **17**, 1571 (1996).

58. K. L. Schuchardt, B. T. Didier, T. Elsethagen, L. Sun, V. Gurumoorthi, J. Chase, J. Li, and T. L. Windus, J. Chem. Inf. Mod. **47**, 1045 (2007).

59. B. Mennucci, E. Cances, and J. Tomasi, J. Phys. Chem. B **101**, 10506 (1997).

60. E. Cancès and B. Mennucci, J. Math. Chem. **23**, 309 (1998).

61. G. Scalmani and M. J. Frisch, J. Chem. Phys. **132**, 114110 (2010).

62. S. P. A. Sauer, V. Špirko, I. Paidarová, and J. Oddershede, Chem. Phys. **184**, 1 (1994).

63. S. P. A. Sauer, V. Špirko, I. Paidarová, and W. P. Kraemer, Chem. Phys. **214**, 91 (1997).

64. R. D. Wigglesworth, W. T. Raynes, J. Oddershede, and S. P. A. Sauer, Mol. Phys. **96**, 1595 (1999).

65. R. D. Wigglesworth, W. T. Raynes, S. Kirpekar, J. Oddershede, and S. P. A. Sauer, J. Chem. Phys. **112**, 736 (2000).

66. T. A. Ruden and K. Ruud, in *Calculation of NMR and EPR Parameters: Theory and Applications*, edited by M. Kaupp, V. G. Malkin, and M. Bühl (Wiley-VCH, Weinheim, 2004), chap. 10, pp. 153–173.

67. R. Faber, A. Buczek, T. Kupka, and S. P. A. Sauer, Mol. Phys. **115**, 144 (2017).

68. R. Faber, J. Kaminsky, and S. P. A. Sauer, in *Gas Phase NMR*, edited by K. Jackowski and M. Jaszunski (Royal Society of Chemistry, London, 2016), chap. 7, pp. 218–266.