Addressing anharmonic effects with density-fitted multicomponent density functional theory

Lukas Hasecke,* Maximilian Breitenbach, Martí Gimferrer, Rainer Oswald, and Ricardo A. Mata*

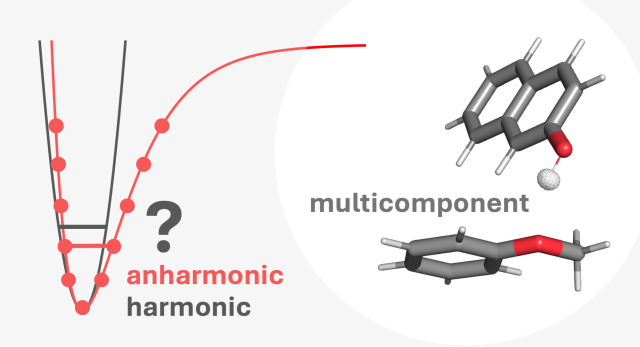
Institute of Physical Chemistry, University of Göttingen, Tammannstrasse 6, 37077, Göttingen, Germany

E-mail: lhaseck@gwdg.de; rmata@gwdg.de

Abstract

In this contribution we present a new density-fitted multicomponent density functional theory implementation and assess its use for the calculation of anharmonic zero-point energies. Four challenging cases of molecular aggregates are reviewed: deprotonated formic acid trimer, diphenyl ether-*tert*-butyl alcohol conformers, anisole/methanol and anisole/2-naphtol dimers. These are all cases where a mismatch between the low-temperature computationally predicted minimum and the experimentally determined structure was observed. Through the use of nuclear-electronic orbital energies in the thermodynamic correction, the correct energetic ordering is recovered. For the smallest system, we compare our results to vibrational perturbation theory anharmonically corrected zero-point energy, with perfect agreement for the lower-lying conformers. The performance of the newly developed code and the density fitting errors are also analysed. Overall, the new implementation shows a very good scaling with system size and the density fitting approximations exhibit a negligible impact.

TOC graphic



Albeit being introduced for several dozens of years,[1–4] multicomponent methods have yet to find widespread usage and an established position in the computational chemist toolbox. These are a unique class of quantum chemical methods, whereby electrons and a selected set of other nuclei (usually protons) are handled on the same footing. From this point onward, and for ease of discussion, we will restrict ourselves to the treatment of protons. The multicomponent methodology entails a partial lifting of the Born-Oppenheimer approximation, with selected protons being described through a wave function and coupling with the electronic solution, either through a mean-field or a more advanced description. One of the most popular denominations for the methods is the nuclear-electronic orbital (NEO) theory,[5] which we also use to denominate our implementations.

When it comes to practicality, density functional theory (DFT) is clearly the framework of choice for several reasons. With the simplicity of a mean-field theory one is still able to handle correlation explicitly through parameterised exchange and correlation functionals. In the case of NEO-DFT,[6–8] this involves solving the Kohn-Sham equations for two different effective potentials

$$\nu^{p}_{\text{eff}}(\mathbf{r}_{1'}) = \sum_{A}^{N_\text{c}} \frac{Z_A}{r_{1'A}} - \sum_{j}^{N_\text{e}} \int \frac{|\phi_j(\mathbf{r}_1)|^2}{r_{11'}} \mathrm{d}\mathbf{r}_1 + \int \frac{\rho^p(\mathbf{r}_{1'})}{r_{1'2'}} \mathrm{d}\mathbf{r}_{2'} + \nu^{ep}_c(\mathbf{r}_{1'}), \quad (1)$$

$$\nu^{e}_{\text{eff}}(\mathbf{r}_{1}) = -\sum_{A}^{N_\text{c}} \frac{Z_A}{r_{1A}} - \sum_{j'}^{N_\text{p}} \int \frac{|\phi_{j'}(\mathbf{r}_{1'})|^2}{r_{11'}} \mathrm{d}\mathbf{r}_{1'} + \int \frac{\rho^e(\mathbf{r}_2)}{r_{12}} \mathrm{d}\mathbf{r}_2 + \nu^{ee}_{xc}(\mathbf{r}_1) + \nu^{ep}_c(\mathbf{r}_1), \quad (2)$$

whereby the indices A stand for N_c classical nuclei, j for the N_e electrons and the dashed indices for the N_p quantum protons. The potentials and densities are marked with e and p accordingly. In the case of the effective potential ν^p_{eff} felt by protons, one has the interaction with classical nuclei, electrons, other quantum protons and the electron-proton (ν^{ep}_c) correlation potential. The electrons and quantum protons couple through the Coulomb interaction and electron-proton correlation. The electron-electron correlation bears the same form as in Born-Oppenheimer calculations.

As one can observe in the equations above, one needs only to provide a functional for electron-electron exchange-correlation (ν^{ee}_{xc}) and for electron-proton correlation (ν^{ep}_c). This is because both proton-proton exchange and correlation are negligible for chemically relevant systems. In our implementation, in order to exclude self-interaction, the diagonal elements of the Hartree-Fock exchange term are added to the protonic effective potential.[9] In comparison to standard electronic structure theory, this only leaves the definition of the electron-proton correlation functional unresolved. It should be noted that since one is dealing with fermion-fermion correlation, any correlation functional form derived from first-principles should be directly applicable, even if it was designed with electrons in mind. In the context of this work, however, we will stray away from this discussion and restrict ourselves to the use of one of the most widely used functionals, the epc-17.2 functional proposed by Hammes-Schiffer and coworkers.[10] The expression is based on the Colle-Salvetti formulae[11] and is provided as

$$E[\rho^e(\mathbf{R}), \rho^p(\mathbf{R})] = \int \frac{-\rho^e(\mathbf{R})\rho^p(\mathbf{R})}{a - b\sqrt{\rho^e(\mathbf{R})\rho^p(\mathbf{R})} + c\rho^e(\mathbf{R})\rho^p(\mathbf{R})} \mathrm{d}\mathbf{R}. \quad (3)$$

The functions $\rho^e(\mathbf{R})$ and $\rho^p(\mathbf{R})$ represent the electronic and nuclear densities at the center of mass of the two particles denoted by $\mathbf{R}$. Due to the significantly larger mass of the proton, $\mathbf{R}$ can be approximated by the proton's position.[9] The parameters $a = 2.35$, $b = 2.40$ and $c = 6.6$ are chosen according to the functional formulation in Ref [10]. We have used the same development framework in Molpro[12] as for our previously published local density-fitted NEO-HF program.[13] The most relevant feature in the case of NEO-DFT is that density fitting approximations are used for the electron-proton Coulomb coupling. The respective densities are approximated by a linear combination of auxiliary density fitting functions

$$\rho_{\mu\nu}(r) \approx \tilde{\rho}_{\mu\nu}(r) = \sum_A d^{\mu\nu}_A \chi_A(r). \quad (4)$$

The indices μ and ν refer in general basis functions, either electronic or nuclear. The fitting coefficients are determined through a so-called robust fitting described in Ref. [13]. The main objective of this work is to demonstrate how this computationally efficient implementation of NEO-DFT can be applied routinely to estimate anharmonic effects in the zero-point vibrational energy of hydrogen-bonded molecular clusters. Inherently including nuclear quantum effects (NQEs) like anharmonic

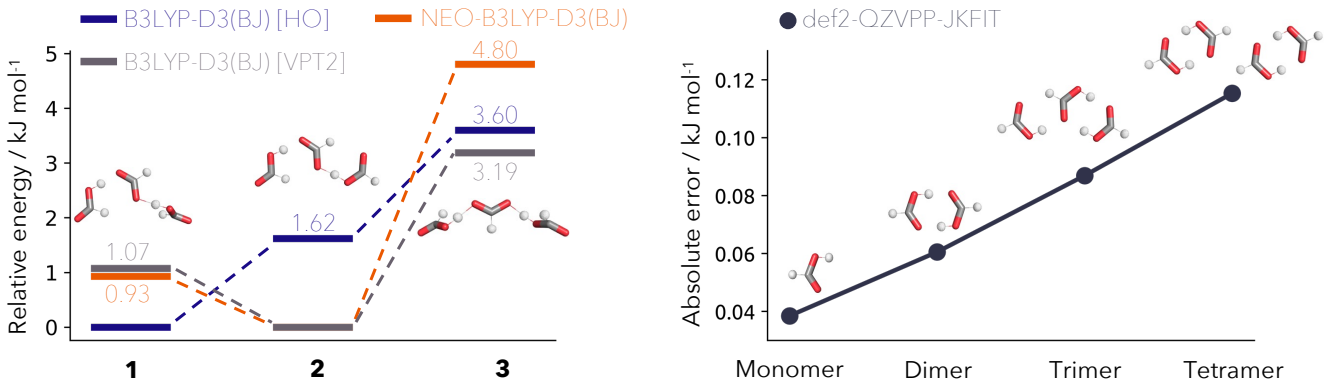


Figure 1: Left: Relative energies of the deprotonated formic acid trimer isomers computed with regular B3LYP-D3(BJ) (blue), NEO-B3LYP-D3(BJ) (orange) and anharmonically corrected via vibrational perturbation theory (gray) utilizing the def2-TZVPP electronic and PB4-F2 nuclear basis set. Right: Absolute error of the density fitted NEO-B3LYP-D3(BJ) method computed with the def2-QZVPP-JKFIT density fitting basis set in comparison to the regular NEO-B3LYP-D3(BJ) results for size increasing formic acid clusters utilizing the def2-TZVP electronic and PB4-F2 nuclear basis set.

zero-point vibrational energy and delocalization within multicomponent DFT carries several advantages.[7,8] These will be demonstrated on challenging examples whereby regular Born-Oppenheimer DFT is either providing inaccurate or mismatching results. For our study we chose the B3LYP-D3(BJ) electronic functional, the most commonly used functional to date for organic compounds.[14,15] All of the experiments later mentioned in the text approach the very-low temperature limit and near-vaccuum conditions (or matrix when explicitly noted) so we will always be comparing the latter to 0 K computed enthalpies.

The first example is the deprotonated formic acid trimer which was thoroughly analysed via infrared action spectroscopy in helium nanodroplets by Taccone *et al.*[16] Three conformers of interest were identified (see Fig. 1). One finds a mismatch between the prevalent structure observed in experiment and the computed global minimum. The potential reasons for this mismatch include the harmonically computed zero-point vibrational energy, as well as potential shortcomings of the experiment, namely kinetics and dynamics of the cluster formation and solvent effects.[16,17] In the closely related proton-bound formate dimer system, constrained NEO molecular dynamics have already demonstrated the usefulness of multicomponent calculations in such strongly-bound hydrogen-bond systems.[18] In Fig. 1 we show the resulting energetic orderings of the three lowest isomers obtained with Born-Oppenheimer based harmonically corrected B3LYP-D3(BJ), anharmonically corrected via vibrational perturbation theory (VPT2) and finally multicomponent NEO-B3LYP-D3(BJ). The harmonic B3LYP-D3(BJ) results are in agreement with the ordering reported by Taccone *et al.*[16] One should note that in their study not only density functional based but also wave function methods, MP2 and CCSD(T), as well as double-hybrid approaches result in the same energetic ordering.[16] The latter is in disagreement with both VPT2 and NEO results. Instead of isomer **1**, isomer **2** becomes the global minimum by 1.07 and 0.93 kJ mol^{-1} for VPT2 and NEO, respectively. These are in agreement with the experimental exclusive observation of isomer **2**. The main disagreement between the VPT2 and the NEO value is found for isomer **3**. NEO-B3LYP-D3(BJ) further destabilizes isomer **3** by 1.20 kJ mol^{-1} compared to the harmonic B3LYP-D3(BJ) result. The VPT2 correction actually stabilizes isomer **3** by 0.41 kJ mol^{-1}. Reviewing the VPT2 calculations it is worth noting that extremely tight criteria for optimisation (and respectively for the DFT numerical grids) have to be used. Otherwise one

risks variations in the relative zero-point vibrational energies of a few kJ/mol. But even with very tight thresholds VPT2 struggles to describe anharmonic O-H bonds, since it is based on a quartic force field built with local information.[19] This is particularly serious for isomer **3**, where the combined symmetric stretching mode of the two bridging protons is red shifted from its harmonic value by $733\,\mathrm{cm}^{-1}$ ($2535\,\mathrm{cm}^{-1}$ vs $1802\,\mathrm{cm}^{-1}$) and exhibits a low overlap with the harmonic mode. This could be the reason for the discrepancy observed in the last isomer.

We now turn to the potential numerical errors introduced by the density fitting approximations used.[12,20] In order to verify the validity of our results we have to benchmark the errors on absolute and relative energies. Therefore, we first employ size increasing clusters of the formic acid molecule ranging from one up to a cluster of four molecules. The obtained results are displayed in Fig. 1. The absolute error introduced by the density fitting scales very low in regards to the overall system size. For the formic acid monomer the absolute error is only $0.04\,\mathrm{kJ\,mol}^{-1}$ and increases only slightly up to $0.12\,\mathrm{kJ\,mol}^{-1}$ for the formic acid tetramer. The absolute error of the trimer, which is slightly bigger than its deprotonated counterpart, is only $0.09\,\mathrm{kJ\,mol}^{-1}$. In order to also assess the impact of the density fitting on the relative energies we analysed the differences in the energies for density-fitted B3LYP-D3(BJ) and regular B3LYP-D3(BJ) results. The obtained energies are shown in Tab. S 1. The overall root mean square deviation between the relative energies is only $0.001\,\mathrm{kJ\,mol}^{-1}$. Such a minor error on relative energies is widely expected.[21–24] In general, the error introduced by density fitting approximations, especially for relative energies, seems negligible.

The next system we would like to revisit with NEO-DFT is the diphenyl ether–*tert*-butyl alcohol complex. The complex was investigated with a multi-spectroscopic approach by Bernhard *et al.*[25] This included FTIR spectroscopy, IR/UV spectroscopy and chirp pulse Fourier transform microwave spectroscopy together with a broad mixture of theoretical approaches going from B3LYP-D3(BJ), over MP2 and CC2 calculations to symmetry adapted perturbation theory.[25] Experimentally the results point in the direction that the OH-O bound dimer is slightly more stable than the OH-π bound isomer. This was derived based on almost similar abundance resulting from the FTIR spectra from the helium expansion, the higher abundance of the OH-O isomer in the mass- and isomer-selective IR/R2PI spectra in neon expansion supported by the broadband rotational spectra in helium and neon.[25] In general the experimental preference towards the OH-O bound dimer was estimated to be in the energy range of $0-1\,\mathrm{kJ\,mol}^{-1}$. Theoretical results however vary largely for the methods employed in their analysis.[25] We recomputed the most stable isomers, predicted by the previous work, with regular B3LYP-D3(BJ) and NEO-B3LYP-D3(BJ). The obtained results are shown in Fig. 2. In agreement with prior B3LYP-D3(BJ) results the most stable isomer is predicted to be OH-π bound. It is a rather small energetic gap to the OH-O bound isomer with only $0.23\,\mathrm{kJ\,mol}^{-1}$, which would be in agreement with the experimental observation of both species. Moreover, the second determined OH-π' structure is energetically $0.33\,\mathrm{kJ\,mol}^{-1}$ higher than the OH-π structure which again is in agreement with experiment. By rotational spectroscopy the OH-π structure was assigned to be the observed species instead of

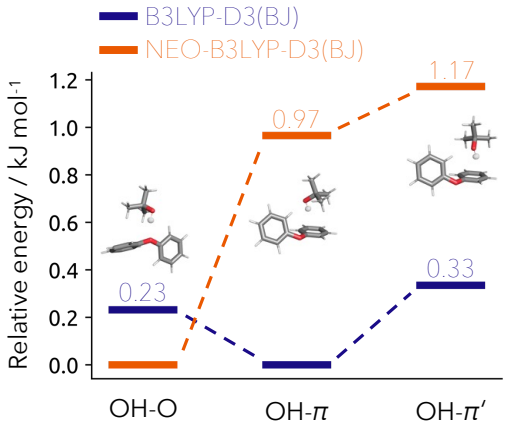


Figure 2: Relative energies of the diphenyl ether-*tert*-butyl alcohol isomers computed with regular B3LYP-D3(BJ) (blue) and NEO-B3LYP-D3(BJ) (orange) utilizing the def2-TZVP electronic and PB4-F2 nuclear basis set.

OH-π'.[25] However, B3LYP-D3(BJ) predicts the OH-π bound dimer to be energetically preferred over the OH-O bound dimer which is against the experimental conclusions. By employing NEO-B3LYP-D3(BJ) the ordering of the PES changes. The global minimum obtained from the multicomponent PES is the OH-O bound dimer, whereas the OH-π bound dimer is energetically 0.97 kJ mol^{-1} higher in energy. Both observations are in agreement with the experimentally derived balance of the OH-O and OH-π bound dimer. In addition, the OH-π' structure is energetically 0.20 kJ mol^{-1} higher as the OH-π structure, which agrees with the rotational spectroscopy experiment. Overall, the results obtained by NEO-B3LYP-D3(BJ) agree in all points with the experimental observations with minimal added computational effort.

We now turn to the last two reference systems. The employed examples are dimers formed by anisole with either methanol or 2-naphthol. Both dimers have already been extensively analysed experimentally and theoretically.[26,27] We start our discussion with the anisole methanol dimer. The experimental analysis of this system is based on FTIR spectroscopy from a cold supersonic jet expansion in helium.[26] They measured a 20 times lower abundance of the OH-π isomer than the preferred OH-O isomer. As a result they estimate the OH-π bound dimer to be at least 1 kJ mol^{-1} less stable than the OH-O dimer. The theoretical methods applied to this system give different results. In case of MP2 the OH-π dimer is clearly favoured by 0.6 kJ mol^{-1}, B3LYP-D3(BJ) favours the OH-O dimer by 0.7 kJ mol^{-1}, the double-hybrid B2PLYP favours the OH-O dimer by 1.1 kJ mol^{-1} and CCSD(T) indicates an almost isoenergetic balance between the two.[26] In general they found, that the balance could be further shifted towards the OH-O bound dimer by 0.2 kJ mol^{-1}, if the aug-cc-pVTZ basis set instead of the def2-TZVP basis set is employed for the geometry optimization and energy computation.[26,28,29] However, the dependence of the initial geometry is dramatically reduced by employing multicomponent methods since the quantum particle will instantaneously adapt to its preferred position. This was already demonstrated in a thorough benchmark for the hydrogen bound methanol complexes with different furan derivatives.[30] Therefore, we first compare the B3LYP-D3(BJ), NEO-B3LYP-D3(BJ) as well as NEO(MP2)-PNO-LCCSD(T)-F12[31] results on the basis of the B3LYP-D3(BJ)/def2-TZVP optimised structures. The obtained results are displayed in Fig. 3. The regular B3LYP-D3(BJ) method leads to a preference towards the OH-O structure with 0.7 kJ mol^{-1} compared to the OH-π bound dimer. This is in agreement with the results obtained by Heger *et al.*[26] Utilizing NEO-B3LYP-D3(BJ) and NEO(MP2)-PNO-LCCSD(T)-F12 instead of regular B3LYP-D3(BJ) carries several advantages as mentioned before. First of all the dependency of the initial geometry is reduced, allowing for computational efficient optimization with small basis sets.[30,31] As a result, the NEO-B3LYP-D3(BJ) single point calculation utilizing the def2-TZVP basis set already leads to experimental comparable results by favoring the OH-O bound dimer by 2.99 kJ mol^{-1}. Compared to the NEO(MP2)-PNO-LCCSD(T)-F12 method, which includes explicit correlation in order to achieve almost complete basis set results, the NEO-B3LYP-D3(BJ) method performs very similarly. Given an absolute energy difference of only 0.12 kJ mol^{-1} between both methods, the NEO-B3LYP-D3(BJ) seems to perform very well on this system. Moreover, the NEO-B3LYP-D3(BJ) exhibits favourable computational effort to achieve the result compared to the higher level approach. The NEO(MP2)-PNO-LCCSD(T)-F12 was running with a total CPU time of 193 minutes to compute both systems, compared to 12 minutes for the NEO-B3LYP-D3(BJ) calculations.

In order to elucidate the basis set dependence of the DFT based results, we computed the energy balance of the anisole-methanol dimer for different basis sets as shown in Tab. S 2. Thereby, the geometry optimization, single point energy calculation and zero-point vibrational energy correction are computed all at the same level of theory. Our harmonically corrected results are in agreement with the values

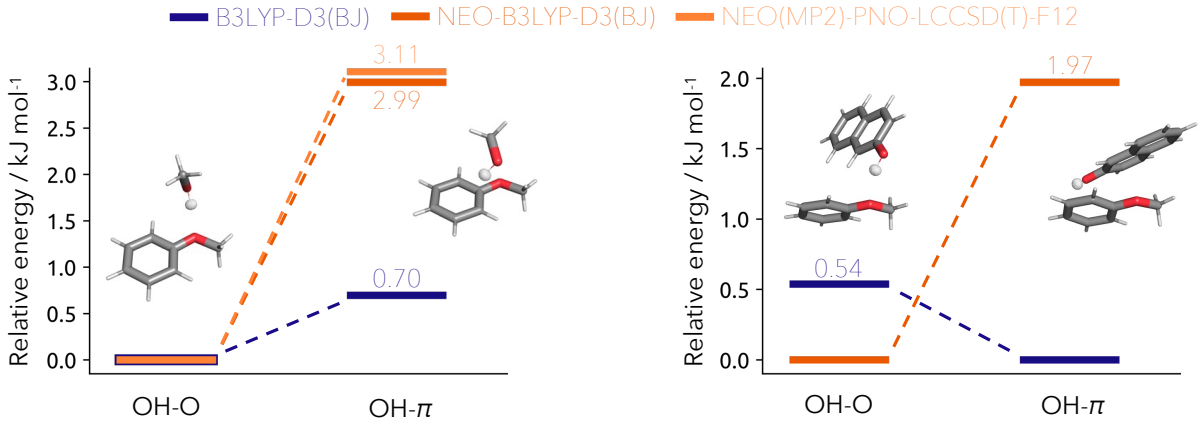


Figure 3: Left: Relative energies of the anisole-methanol dimer isomers computed with regular B3LYP-D3(BJ) (blue) and NEO-B3LYP-D3(BJ) (orange) utilizing the def2-TZVP electronic and PB4-F2 nuclear basis set together with the NEO(MP2)-PNO-LCCSD(T)-F12 (light orange) method utilizing the cc-pVTZ-F12 electronic and PB4-F2 nuclear basis set. Right: Relative energies of the anisol 2-naphthol dimer isomers computed with regular B3LYP-D3(BJ) (blue) and NEO-B3LYP-D3(BJ) (orange) utilizing the def2-TZVP electronic and PB4-F2 nuclear basis set.

provided by Heger *et al.*[26] With increasing basis set size, the OH-π bound dimer relative energy is estimated at about 1 kJ mol^{-1}. The experimental observations would be more in line with an energy difference of 2 kJ mol^{-1} or more.[26] Our multicomponent approach placed the relative energy at about 3.5 kJ mol^{-1}.

We would like to emphasize the advantages of NEO-DFT also for the next test system, the dimer formed between anisole and 2-naphthol. Experimentally a combination of jet-cooled FTIR spectroscopy and laser-induced fluorescence spectroscopy, as well as resonance-enhanced two-photon UV ionisation spectroscopy was employed by Nejad *et al.*[27] They observed only the OH-O bound dimer, whereas the OH-π bound dimer remains experimental elusive.[27] The theoretical methods applied also provide diverse results from B3LYP-D3(BJ) favoring the OH-π structure by 0.7 kJ mol^{-1} and 0.5 kJ mol^{-1} for the def2-TZVP and def2-QZVP basis sets and only twist the preference slightly to the OH-O structure with 0.1 kJ mol^{-1} and 0.2 kJ mol^{-1} for the def2-TZVP and def2-QZVP basis sets if three-body dispersion corrections are included.[27] However, neither of those results would energetically explain the complete elusiveness of the OH-π dimer. Moreover, unscaled MP2 variants also favour the OH-π structure. By employing SCS-MP2 or PNO-LCCSD(T)-F12b the OH-O structure is favoured.[27] PNO-LCCSD(T)-F12b/cc-pVTZ-F12 should be within a kJ mol^{-1} of the complete basis set limit of CCSD(T),[22,23,32] the current gold standard of quantum chemistry. In this case, the OH-O structure is about 3 kJ mol^{-1} lower in energy. Overall, this provides an excellent benchmark for the NEO-B3LYP-D3(BJ) method. The results obtained for the anisol 2-naphthol dimer with regular B3LYP-D3(BJ) and NEO-B3LYP-D3(BJ) are displayed in Fig. 3. B3LYP-D3(BJ) prefers the OH-π bound dimer by 0.54 kJ mol^{-1} which is in agreement with the results from Nejad *et al.*.[27] However, by employing NEO-B3LYP-D3(BJ) the PES changed noticeably to a strong preference of the OH-O bound dimer which is energetically favoured by 1.97 kJ mol^{-1}. Compared to the PNO-LCCSD(T)-F12b and the experimental conclusion, the obtained results with NEO-B3LYP-D3(BJ) are in line with both. However, it should be noted that the NEO-B3LYP-D3(BJ) results are obtained at much lower computational costs in comparison to the PNO-LCCSD(T)-F12b results.

In conclusion, the NEO-B3LYP-D3(BJ) method recovers in all tested cases a PES which is in agreement with experimental observations. It thus provides a feasible computational method to obtain (partially) corrected PESs

which directly include NQEs like anharmonic zero-point energies and captures the nuclear delocalization. This is a step forward in providing the right answer for the right reasons for challenging cases to regular DFT calculations. Therefore, it is particularly useful for the analysis of experiments targeting cold molecular clusters. Questions surrounding the kinetics and dynamics of cluster formation and solvent effects still remain elusive. To tackle those effects, vibrational NEO based analysis as well as NEO molecular dynamics will be employed in future studies.[33–35]

Computational Methods

All NEO calculations have been carried out with Molpro,[12] employing the B3LYP functional including the D3 dispersion correction with Becke-Johnson damping.[14,36] The (local) density fitted version of the NEO-DFT module integrated in Molpro is an extension of our previously presented (local) density fitted NEO restricted HF implementation.[13] Thereby, we employ (local) density fitting within the electronic subsystem and for the Coulomb coupling between quantum mechanical treated nuclei and electrons. The subsystem of the quantum nuclei is treated by an integral-direct implementation.[13,30] Overall, the general implementation together with accuracy and performance of density fitting within multicomponent DFT was previously discussed by Mejía-Rodríguez *et al.*[20] For the quantum mechanical nuclei the PB4-F2 nuclear basis set together with the even tempered 10s10p10d10f fitting basis set with exponents ranging from $2\sqrt{2}$ to 64 are employed.[37] Both NEO and regular DFT calculations are carried out employing the def2-TZVP, or def2-TZVPP in the case of the deprotonated formic acid trimer, basis sets with the def2-QZVPP-JKFIT density fitting basis set.[38,39] A threshold of 10^{-7} a.u. for the energy difference within the electronic and nuclear SCF computations, the difference in the density between iterations and the gradient of the respective nuclear and electronic subiterations. The overall energy difference in the NEO-DFT iterations was set to a threshold of 10^{-6} Hartree. All Molpro computations employ the direct inversion in the iterative subspace starting after the first iteration with a maximum of 10 Fock matrices as basis to extrapolate.[40,41] In general the standard grid 3 is employed for the computations, whereas the formic acid clusters are computed with the standard grid 2.[42] The electron-proton correlation is computed with the epc-17.2 functional.[10] In order to assess the error of the density fitting, reference calculations for the formic acid clusters with regular NEO-DFT are carried out with Q-Chem 6.2.[43] Thereby, the standard grid 2, a threshold of 10^{-8} Hartree for the energy and the geometric direct minimization algorithm were employed.[42,44] For those systems also the threshold within Molpro was raised to 10^{-8} a.u. for the energy difference within the NEO microiterations and 10^{-7} for the overall energy difference. These tighter thresholds were also employed for the NEO-RHF reference wave function of the NEO(MP2)-PNO-LCCSD(T)-F12 method.[30] Those calculations employ the cc-pVTZ-F12 basis set with the cc-pVQZ-JKFIT density fitting basis for the Fock and the exchange matrices as well as the complementary auxiliary basis set for the resolution of the identity and the cc-pVQZ-MP2FIT density fitting basis set.[32,45,46] The F12b energies are obtained with the 3*A(LOC,FIX) ansatz. Moreover, the complementary auxiliary basis set singles correction together with the scaling of the perturbative triples are applied.[21–23,47] All systems are optimized with Gaussian16 employing the def2-TZVP basis set with very tight SCF settings, tight optimization thresholds and a superfine grid utilizing B3LYP-D3(BJ).[48] In the case of the deprotonated formic acid trimer the def2-TZVPP basis set is employed.[38] Corresponding frequency calculations are carried out with the same settings. In order to obtain the zero-point vibrational energies of the systems without the contributions of the quantum mechanical treated protons, the isotope mass of the respective nuclear centers was set to $9.9{\cdot}10^{12}$ a.u., making those centers infinitely heavy. All nuclear densities shown are displayed at a $0.02\,\sigma$ contour level generated with the PyMOL 2.5.2 program.[49]

Acknowledgements

The authors gratefully acknowledge funding of this research by the German Research Foundation (DFG): L.H. via project 389479699/GRK2455, M.B. via project 510228793/SFB1633 (project B01) and M.G. via project 217133147/SFB1073 (project C03).

Supporting Information

Relative energies of the deprotonated formic acid trimers with respect to the energetically most stable isomer **1** Table S1, relative energies of the anisole-methanol OH-π bound isomer with respect to the OH-O bound isomer for different basis sets Table S2 (PDF). Structural information and energies are available free of charge on GRO.data (`https://doi.org/10.25625/2IPVNS`).

Competing interests

The authors declare no competing interest.

References

(1) Thomas, I. L. Protonic Structure of Molecules. I. Ammonia Molecules. *Phys. Rev.* **1969**, *185*, 90–94.

(2) Thomas, I. The protonic structure of methane, ammonia, water, and hydrogen fluoride. *Chemical Physics Letters* **1969**, *3*, 705–706.

(3) Capitani, J. F.; Nalewajski, R. F.; Parr, R. G. Non-Born–Oppenheimer density functional theory of molecular systems. *The Journal of Chemical Physics* **1982**, *76*, 568–573.

(4) Kreibich, T.; Gross, E. K. U. Multicomponent Density-Functional Theory for Electrons and Nuclei. *Phys. Rev. Lett.* **2001**, *86*, 2984–2987.

(5) Pavošević, F.; Culpitt, T.; Hammes-Schiffer, S. Multicomponent Quantum Chemistry: Integrating Electronic and Nuclear Quantum Effects via the Nuclear–Electronic Orbital Method. *Chemical Reviews* **2020**, *120*, 4222–4253, PMID: 32283015.

(6) Pak, M. V.; Chakraborty, A.; Hammes-Schiffer, S. Density Functional Theory Treatment of Electron Correlation in the Nuclear-Electronic Orbital Approach. *The Journal of Physical Chemistry A* **2007**, *111*, 4522–4526.

(7) Chakraborty, A.; Pak, M. V.; Hammes-Schiffer, S. Development of Electron-Proton Density Functionals for Multicomponent Density Functional Theory. *Phys. Rev. Lett.* **2008**, *101*, 153001.

(8) Chakraborty, A.; Pak, M. V.; Hammes-Schiffer, S. Properties of the exact universal functional in multicomponent density functional theory. *The Journal of Chemical Physics* **2009**, *131*, 124115.

(9) Yang, Y.; Brorsen, K. R.; Culpitt, T.; Pak, M. V.; Hammes-Schiffer, S. Development of a practical multicomponent density functional for electron-proton correlation to produce accurate proton densities. *The Journal of Chemical Physics* **2017**, *147*, 114113.

(10) Brorsen, K. R.; Yang, Y.; Hammes-Schiffer, S. Multicomponent Density Functional Theory: Impact of Nuclear Quantum Effects on Proton Affinities and Geometries. *The Journal of Physical Chemistry Letters* **2017**, *8*, 3488–3493.

(11) Colle, R.; Salvetti, O. Approximate calculation of the correlation energy for the closed shells. *Theoretica Chimica Acta* **1975**, *37*, 329–334.

(12) Werner, H.-J.; Knowles, P. J.; Celani, P.; Györffy, W.; Hesselmann, A.; Kats, D.; Knizia, G.; Köhn, A.; Korona, T.; Kreplin, D. et al. MOLPRO, 2024.1 , a package of ab initio programs. see https://www.molpro.net.

(13) Hasecke, L.; Mata, R. A. Nuclear Quantum Effects Made Accessible: Local Density Fitting in Multicomponent Methods. *Journal of Chemical Theory and Computation* **2023**, *19*, 8223–8233.

(14) Tirado-Rives, J.; Jorgensen, W. L. Performance of B3LYP Density Functional Methods for a Large Set of Organic Molecules. *Journal of Chemical Theory and Computation* **2008**, *4*, 297–306, PMID: 26620661.

(15) Tirado-Rives, J.; Jorgensen, W. L. Performance of B3LYP Density Functional Methods for a Large Set of Organic Molecules. *Journal of Chemical Theory and Computation* **2008**, *4*, 297–306.

(16) Taccone, M. I.; Thomas, D. A.; Ober, K.; Gewinner, S.; Schöllkopf, W.; Meijer, G.; von Helden, G. Infrared action spectroscopy of the deprotonated formic acid trimer, trapped in helium nanodroplets. *Physical Chemistry Chemical Physics* **2023**, *25*, 10907–10916.

(17) Meyer, K. A. E.; Davies, J. A.; Ellis, A. M. Shifting formic acid dimers into perspective: vibrational scrutiny in helium nanodroplets. *Physical Chemistry Chemical Physics* **2020**, *22*, 9637–9646.

(18) Zhang, Y.; Xu, X.; Yang, N.; Chen, Z.; Yang, Y. Describing proton transfer modes in shared proton systems with constrained nuclear–electronic orbital methods. *The Journal of Chemical Physics* **2023**, *158*.

(19) Fortenberry, R. C.; Yu, Q.; Mancini, J. S.; Bowman, J. M.; Lee, T. J.; Crawford, T. D.; Klemperer, W. F.; Francisco, J. S. Communication: Spectroscopic consequences of proton delocalization in OCHCO+. *The Journal of Chemical Physics* **2015**, *143*.

(20) Mejía-Rodríguez, D.; de la Lande, A. Multicomponent density functional theory with density fitting. *The Journal of Chemical Physics* **2019**, *150*.

(21) Ma, Q.; Werner, H.-J. Scalable Electron Correlation Methods. 2. Parallel PNO-LMP2-F12 with Near Linear Scaling in the Molecular Size. *Journal of Chemical Theory and Computation* **2015**, *11*, 5291–5304.

(22) Ma, Q.; Werner, H. Explicitly correlated local coupled-cluster methods using pair natural orbitals. *WIREs Computational Molecular Science* **2018**, *8*.

(23) Ma, Q.; Werner, H.-J. Accurate Intermolecular Interaction Energies Using Explicitly Correlated Local Coupled Cluster Methods [PNO-LCCSD(T)-F12]. *Journal of Chemical Theory and Computation* **2019**, *15*, 1044–1052.

(24) Werner, H.-J.; Knizia, G.; Krause, C.; Schwilk, M.; Dornbach, M. Scalable Electron Correlation Methods I.: PNO-LMP2 with Linear Scaling in the Molecular Size and Near-Inverse-Linear Scaling in the Number of Processors. *Journal of Chemical Theory and Computation* **2015**, *11*, 484–507.

(25) Bernhard, D.; Dietrich, F.; Fatima, M.; Perez, C.; Poblotzki, A.; Jansen, G.; Suhm, M. A.; Schnell, M.; Gerhards, M. Multi-spectroscopic and theoretical analyses on the diphenyl ether–tert-butyl alcohol complex in the electronic ground and electronically excited state. *Physical Chemistry Chemical Physics* **2017**, *19*, 18076–18088.

(26) Heger, M.; Altnöder, J.; Poblotzki, A.; Suhm, M. A. To π or not to π – how does methanol dock onto anisole? *Physical Chemistry Chemical Physics* **2015**, *17*, 13045–13052.

(27) Nejad, A.; Pérez Mellor, A. F.; Lange, M.; Alata, I.; Zehnacker, A.; Suhm, M. A. Subtle hydrogen bond preference and dual Franck–Condon activity – the interesting pairing of 2-naphthol with anisole. *Physical Chemistry Chemical Physics* **2023**, *25*, 10427–10439.

(28) Dunning Jr, T. H.; Peterson, K. A.; Wilson, A. K. Gaussian basis sets for use in correlated molecular calculations. X. The atoms aluminum through argon revisited. *The Journal of Chemical Physics* **2001**, *114*, 9244–9253.

(29) Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.

(30) Hasecke, L.; Mata, R. A. Local electronic correlation in multicomponent Møller-Plesset perturbation theory. **2024**,

(31) Hasecke, L.; Mata, R. A. Optimization of Quantum Nuclei Positions with the Adaptive Nuclear-Electronic Orbital Approach. *The Journal of Physical Chemistry A* **2024**, *128*, 3205–3211.

(32) Peterson, K. A.; Adler, T. B.; Werner, H.-J. Systematically convergent basis sets for explicitly correlated wavefunctions: The atoms H, He, B–Ne, and Al–Ar. *The Journal of Chemical Physics* **2008**, *128*.

(33) Culpitt, T.; Yang, Y.; Schneider, P. E.; Pavošević, F.; Hammes-Schiffer, S. Molecular Vibrational Frequencies with Multiple Quantum Protons within the Nuclear-Electronic Orbital Framework. *Journal of Chemical Theory and Computation* **2019**, *15*, 6840–6849.

(34) Chow, M.; Li, T. E.; Hammes-Schiffer, S. Nuclear–Electronic Orbital Quantum Mechanical/Molecular Mechanical Real-Time Dynamics. *The Journal of Physical Chemistry Letters* **2023**, *14*, 9556–9562.

(35) Xu, X.; Chen, Z.; Yang, Y. Molecular Dynamics with Constrained Nuclear Electronic Orbital Density Functional Theory: Accurate Vibrational Spectra from Efficient Incorporation of Nuclear Quantum Effects. *Journal of the American Chemical Society* **2022**, *144*, 4039–4046, PMID: 35196860.

(36) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *Journal of Computational Chemistry* **2011**, *32*, 1456–1465.

(37) Yu, Q.; Pavošević, F.; Hammes-Schiffer, S. Development of nuclear basis sets for multicomponent quantum chemistry methods. *The Journal of Chemical Physics* **2020**, *152*, 244123.

(38) Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Physical Chemistry Chemical Physics* **2005**, *7*, 3297.

(39) Weigend, F. Hartree–Fock exchange fitting basis sets for H to Rn †. *Journal of Computational Chemistry* **2007**, *29*, 167–175.

(40) Pulay, P. Convergence acceleration of iterative sequences. the case of scf iteration. *Chemical Physics Letters* **1980**, *73*, 393–398.

(41) Pulay, P. Improved SCF convergence acceleration. *Journal of Computational Chemistry* **1982**, *3*, 556–560.

(42) Dasgupta, S.; Herbert, J. M. Standard grids for high-precision integration of modern density functionals: SG-2 and SG-3. *Journal of Computational Chemistry* **2017**, *38*, 869–882.

(43) Epifanovsky, E.; Gilbert, A. T. B.; Feng, X.; Lee, J.; Mao, Y.; Mardirossian, N.; Pokhilko, P.; White, A. F.; Coons, M. P.; Dempwolff, A. L. et al. Software for the frontiers of quantum chemistry: An overview of developments in the Q-Chem 5 package. *The Journal of Chemical Physics* **2021**, *155*, 084801.

(44) Van Voorhis, T.; Head-Gordon, M. A geometric approach to direct minimization. *Molecular Physics* **2002**, *100*, 1713–1721.

(45) Weigend, F. A fully direct RI-HF algorithm: Implementation, optimised auxiliary basis sets, demonstration of accuracy and efficiency. *Phys. Chem. Chem. Phys.* **2002**, *4*, 4285–4291.

(46) Weigend, F.; Köhn, A.; Hättig, C. Efficient use of the correlation consistent basis sets in resolution of the identity MP2 calculations. *The Journal of Chemical Physics* **2002**, *116*, 3175–3183.

(47) Knizia, G.; Adler, T. B.; Werner, H.-J. Simplified CCSD(T)-F12 methods: Theory and benchmarks. *The Journal of Chemical Physics* **2009**, *130*.

(48) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H. et al. Gaussian 16 Revision A.03. 2016; Gaussian Inc. Wallingford CT.

(49) The PyMOL Molecular Graphics System, Version 2.5.2 Schrödinger, LLC.