Relative energetics of acetyl-histidine protomers with and without Zn^{2+} and a benchmark of energy methods

Markus Schneider[†] and Carsten Baldauf[*,†,‡]

†*Fritz-Haber-Institut der Max-Planck-Gesellschaft, Theory Department, Faradayweg 4-6, D-14195 Berlin, Germany*

‡*Wilhelm-Ostwald-Institut für Physikalische und Theoretische Chemie, Fakultät für Chemie und Mineralogie, Universität Leipzig, Linnéstrasse 2, 04103 Leipzig, Germany*

E-mail: baldauf@fhi-berlin.mpg.de

Abstract

We studied acetylhistidine (AcH), bare or microsolvated with a zinc cation by simulations in isolation. First, a global search for minima of the potential energy surface combining both, empirical and first-principles methods, is performed individually for either one of five possible protonation states. Comparing the most stable structures between tautomeric forms of negatively charged AcH shows a clear preference for conformers with the neutral imidazole ring protonated at the $N_{\epsilon2}$ atom. When adding a zinc cation to the system, the situation is reversed and $N_{\delta1}$-protonated structures are energetically more favorable. Obtained minima structures then served as basis for a benchmark study to examine the goodness of commonly applied levels of theory, *i.e.* force fields, semi-empirical methods, density-functional approximations (DFA), and wavefunction-based methods with respect to high-level coupled-cluster calculations, *i.e.* the DLPNO-CCSD(T) method. All tested force fields and semi-empirical methods show a poor performance in reproducing the energy hierarchies of conformers, in particular of systems involving the zinc cation. Meta-GGA, hybrid, double hybrid DFAs, and the MP2 method are able to describe the energetics of the reference method within "chemical accuracy", *i.e.* with a mean absolute error of less than 1 kcal/mol. Best performance is found for the double hybrid DFA B3LYP+XYG3 with a mean absolute error of 0.7 kcal/mol and a maximum error of 1.8 kcal/mol. While MP2 performs similarly as B3LYP+XYG3, computational costs, i.e. timings, are increased by a factor of 4 in comparison due to the large basis sets required for accurate results.

2

Introduction

Metal cations are essential to life, as approximately one third of the proteins in the human body require a metal cofactor for biological function.[1,2] They often play a crucial role in shaping the three-dimensional structure of proteins and peptides. In their presence, peptides often undergo significant conformational changes - usually imposing structural constraints - that may alter important properties, *e.g.* binding sites, catalytic properties, and biological functions. As an example, it is hypothesized that protein misfolding of Alzheimer's Aβ-amyloid peptides into aggregated senile plaques inside the human brain of Alzheimer patients is promoted by metal ions such as zinc (Zn^{2+}).[3] Zinc ions are furthermore required for the catalytic function of more than 200 enzymes,[4] an example being carbonic anhydrase essential to the process of carbon dioxide regulation.[5]

It is therefore much desirable to have a very good fundamental and detailed theoretical understanding of interactions of metal cations with peptides. The goal of this research is to investigate the energetics of peptides in conjunction with metal cations, with a strong focus on benchmark systems consisting of either a bare acetylhistidine (AcH) or microsolvated with a Zn^{2+} cation. The goodness of commonly applied theoretical levels of theory, *i.e.* force field (FF), semi-empirical quantum chemistry methods, density-functional approximation (DFA), and wavefunction-based methods is being assessed and evaluated with respect to high-level coupled-cluster calculations. The main focus lies thereby on peptide structures in the gas-phase as opposed to solvated systems. This is mainly due to the desire of studying the inner, "undamped" interactions and forces of molecules and their correct description by theoretical models without having to deal with models outside of the core problem, *e.g.* implicit solvent treatment models, that provide an additional non-negligible uncertainty. Furthermore, the chosen system of AcH serves as a benchmark system as it is still computationally feasible, even for high-level methods, yet provides a challenging structure because of the tautomeric form of its neutral imidazole ring.

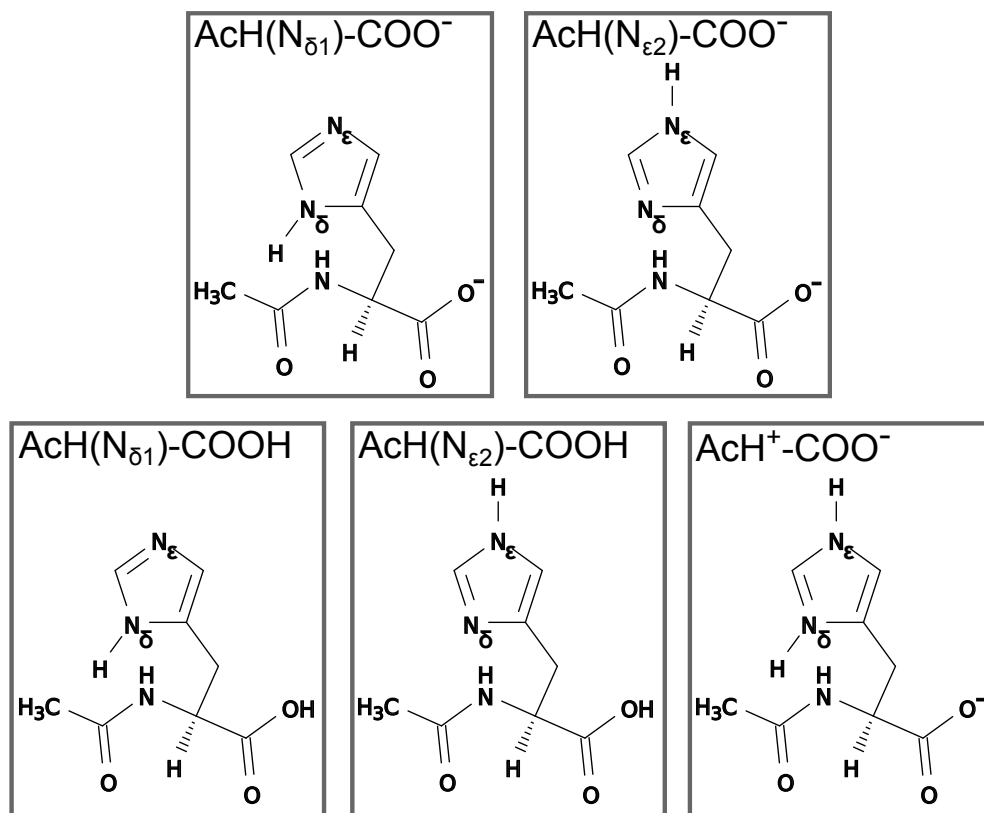Figure 1 shows chemical structures of AcH with the different protonation states in-

Figure 1: Chemical structures of negatively charged AcH (upper row) showing the two equivalent tautomeric forms of the neutral imidazole side chain. For neutral AcH (bottom row), three different protonation states are theoretically possible.

vestigated in this work: Negatively charged AcH (upper row in Figure 1) has two equivalent tautomeric forms of the neutral imidazole side chain. The two forms are labeled $AcH(N_{\delta 1})-COO^-$ and $AcH(N_{\epsilon 2})-COO^-$, meaning that either the $N_{\delta 1}$ or the $N_{\epsilon 2}$ atom is protonated in the imidazole ring. For bare neutral AcH (bottom row in Figure 1), three different protonation states are theoretically possible: Besides the two equivalent tautomeric forms, labeled $AcH(N_{\delta 1})-COOH$ and $AcH(N_{\epsilon 2})-COOH$, that have a neutral carboxyl group at the C-terminus $(-COOH)$, a third form exists, labeled AcH^+-COO^-, which has both the $N_{\delta 1}$ and $N_{\epsilon 2}$ nitrogens of the imidazole protonated but the carboxyl group at the C-terminus deprotonated $(-COO^-)$. As already pointed out, either system is studied bare as well as microsolvated with a Zn^{2+} cation, resulting in ten different systems to be investigated.

Benchmark calculations for small systems containing a zinc cation have been done in the past. Amin and Truhlar set up a benchmark database of Zn coordination compounds with O, S, NH_3, H_2O, OH, SCH_3, and H ligands.[6] Using coupled cluster calculations with augmented polarized triple-ζ basis sets as the reference, 39 density functionals and seven more approximate molecular orbital theories were tested. They found that DFT overall significantly outperformed semi-empirical methods. Best performance was generally found for exchange-correlation functionals containing a portion of Hartree-Fock exchange. Out of the functionals that contained no Hartree-Fock exchange, M06-L[7] displayed the best performance. Similarly, Rayón *et al.* tested the performance of five different functionals against MP2[8,9] and CCSD(T)[10] calculations, with the B3LYP[11,12] functional performing best.[13] Weaver *et al.* predicted nine ZnX complexes (X= Zn, H, O, F_2, S, Cl, Cl_2, CH_3, $(CH_3)_2$) using 14 density functionals, MP2 calculations and the CCSD and CCSD(T) coupled-cluster methods applying correlation consistent triple-ζ basis sets.[14] They found that BLYP,[15,16] B3LYP, MP2, CCSD and CCSD(T) showed poor performances based on accuracy, which for the latter three wavefunction based methods might be caused by a missing complete basis set description[17] or the slow-converging correlation contribution of the zinc electrons that may lead to large and conformation dependent basis set superposition errors (BSSE).[18,19] Gutten

et al. evaluated the performance of the wavefunction-based MP2 method as well as several DFA xc functionals with respect to CCSD(T) using gas-phase complexation energies calculated for five model complexes and four metal ions (Fe^{2+}, Cu^{2+}, Zn^{2+}, Cd^{2+}).[20] Reasonable agreement was found for MP2 with values usually within $1.5\,kcal/mol$ from the reference values, while DFT performed less satisfactory, although the appropriateness of the models may be significantly altered when combining them with advanced solvation models.[21] For certain complexes containing metal-ligand bonds, large errors in the gas-phase complexation energies (with values up to $20\,kcal/mol$) were reported. Performance concerning geometry optimization was found to be satisfactory already using the PBE xc functional on the GGA DFT level. In the benchmark studies by Navrátil *et al.* on activation and reaction energies for four model systems of peptide bond hydrolysis in an ion-free environment and in presence of one and two zinc ions, reasonably good performance was found for several DFAs and MP2 when comparing to CCSD(T)-obtained results.[22] Best performance for calculating activation barriers was achieved when using the B3LYP or the M06-2X xc functionals on the DFA level of theory. An extremely large number of amino acid-cation conformations have been generated and analyzed by Ropo and colleagues: The rigorous structure search provided structure/energy data on equal footing over a significant stretch of biochemical space[23] and allowed for interesting analysis, even linking simple structural properties of the amino acid-cation complexes to acute toxicities.[24] Finally, benchmark evaluations and calibrations of theoretical calculations help in modeling metal-binding sites and studying metal-ion selectivity in proteins.[25–28]

The general approach in this work follows a previous study by us on the accuracy of various energy functions for carbohydrates[29] and is briefly outlined in the following. First, a global search for minima of the potential energy surface (PES) combining both FF and DFA is performed for either one of the ten systems individually. The obtained global minima and energy hierarchies are then discussed and compared for systems of equal overall charge q, *i.e.* $q = 1$ for the upper row in Figure 1 and $q = 0$ for the bottom row in Figure 1. For

the benchmarking studies, a certain set of structures is then selected based on simple energy criteria. While the focus does lie on local minima structures, it is intended to select structures that vary largely in energy and structure in order to intentionally provide "difficulty" for the theoretical methods to be benchmarked. On top of that, all systems carrying the same overall charge q, are benchmarked at once (except for FFs), thus providing even more "challenge" for the methods in question. Finally, across-the-scale total energy calculations for a wide variety of FFs, semi-empirical methods, DFAs, and wavefunction based methods are tested and evaluated against high-level coupled cluster calculations using mean absolute errors (MAEs) and maximum errors (MEs) as a quality measure.

Computational details

Conformational sampling

In order to yield minima structures that serve as a basis for selecting a set of conformers for the benchmarking process, the conformational space needs to be sampled first. To that end, an energy minimum search combining both FF and DFA is laid out. First, a global energy minimum search was performed using a basin-hopping approach within the `TINKER` molecular modeling package.[30,31] The program works on the level of empirical force fields and the 2009 AMOEBA biopolymer force field, labeled AMOEBA-BIO09,[32,33] was applied here, which is for two reasons: First, this polarizable force field provides a much "rougher" potential energy surface than widely used conventional force fields, such as AMBER-99,[34] CHARMM22,[35] or OPLS-AA,[36,37] because it uses atomic charge multipole expansion instead of fixed point charges. The "rougher" the potential energy surface the more minima are found, hence the conformational space is sampled in more detail. For example, depending on the actual studied system, *i.e.* whether the Zn^{2+} cation is present or the protonation state of the imidazole side chain and the carboxyl group, the number of minima found can be up to a factor of six higher when using the AMOEBA-BIO09 force field in comparison

to the OPLS-AA, AMBER-99, and CHARMM22 force fields. Secondly, the AMOEBA-BIO09 FF was the only FF available providing out-of-the-box parameters for the neutral carboxyl group ($-$COOH). Concerning the technical aspect of the basin-hopping search, the `scan` subprogram within `TINKER` has been applied using all automatically found torsional angles, a relative energy window of 100 kcal/mol and an energy similarity criterion of 0.0001 kcal/mol. After having applied the FF driven basin-hopping approach, all found minima were locally refined, *i.e.* the individual structures were being geometrically relaxed, using DFA implemented within the all-electron/full-potential electronic structure code package `FHI-aims`.[38–40] To be more precise, relaxation was accomplished using a trust radius method version of the Broyden-Fletcher-Goldfarb-Shanno (BFGS) optimization algorithm.[41] First, local refinement was done on the PBE+vdW level using `FHI-aims` specific `tier 1` basis sets and `light` settings intended to give reliable energies for screening purposes.[38] This means the PBE[42] generalized gradient approximation (GGA) exchange-correlation (xc) functional has been applied along with the Tkatchenko-Scheffler[43] van der Waals scheme (vdW) for treating long-range dispersion effects. To rule out duplicate structures, a clustering scheme was applied. To be precise, root-mean-square deviations (RMSD) of atomic positions between any two conformers were calculated using `OpenBabel`.[44] Hierarchical clustering was then achieved by applying the Unweighted Pair Group Method with Arithmetic Mean (UPGMA)[45] method implemented in `Python`'s `SciPy`[46] library. Following that, further relaxation was accomplished at the PBE+vdW level using `tier 2` basis sets and `tight` settings that are intended to provide meV-level accurate energy differences,[38] *i.e.* within 0.02 kcal/mol. After clustering, further relaxation was accomplished at the PBE0+MBD level using the same two-step approach as before, *i.e.* using first `tier 1` basis sets and `light` settings, and `tier 2` basis sets and `tight` settings afterwards. As the labeling suggests, the PBE0[47] hybrid xc functional has been applied, augmented by a many-body dispersion (MBD)[48] correction for long-range dispersion treatment.

Levels of theory and energy calculation methods

Wavefunction-based methods

The benchmark calculations in this paper are based on high-level coupled-cluster calculations.[49,50] In particular, the coupled-cluster method including single, double, and perturbative triple excitations, named CCSD(T),[10] is commonly referred to as the "gold standard of quantum chemistry" due to its high accuracy in the complete basis set limit (CBS).[51] However, due to the slow convergence of the electronic correlation energy with basis set size N as well as the technique's $\mathcal{O}(N^7)$-scaling of the computational costs, accurate results that require large enough basis sets are currently not affordable for system sizes treated in this work. Instead, the domain-based local pair natural orbital (DLPNO-)CCSD(T)[52,53] technique serves as the reference method in this work. The DLPNO-CCSD(T) approximation aims to fully exploit locality of the electron correlation and shows a near linear scaling behavior with basis set size N. Calculations have been carried out with the electronic structure program package `ORCA`.[54] Ahlrichs' `def2`[55] basis set family has been used throughout for all wavefunction-based methods. Because heavy elements like Zn^{2+} require a relativistic treatment of all-electron calculations, the 0th order regular approximation (ZORA),[56] implemented in `ORCA` in an approximate way,[57] is used throughout. As the scalar relativistic treatment requires flexible basis sets, this in turn means that `ORCA` automatically provides relativistically recontracted versions[58] of Ahlrichs' `def2` basis set family, labeled `ZORA-def2`. The accuracy of the DLPNO-CCSD(T) method has been tested previously with a series of benchmark sets covering a broad range of quantum chemical applications.[59] An accuracy of $1\,\text{kcal/mol}$ commonly named *chemical accuracy*, could be obtained using `normal` settings. Still, before using the DLPNO-CCSD(T) method with `normal` settings as the reference method in this work, validation has to be done against conventional CCSD(T) calculations for the systems depicted in Figure 1 and using Ahlrichs' relativistically recontracted split valence basis set with added polarization functions, labeled ZORA-def2-SVP.

The other post-Hartree-Fock *ab initio* method in this work to be benchmarked against DLPNO-CCSD(T), is the widely used second-order Møller-Plesset perturbation theory (MP2).[8,9] Calculations have been carried out again with `ORCA` and applying a resolution of identity (RI) approximation.[60]

Energy calculations for both DLPNO-CCSD(T) and MP2 have been performed using Ahlrichs's ZORA-def2-SVP basis set as well as relativistically recontracted valence triple-zeta and quadruple-zeta basis sets with two sets of polarization functions added, labeled ZORA-def2-TZVPP and ZORA-def2-QZVPP, respectively. Extrapolation to the CBS limit has been applied on calculated Hartree-Fock (HF) energies and correlation energies individually. To be more precise, HF energies have been extrapolated using a form proposed by Karton and Martin:[61]

$$E_{HF}^n = E_{HF}^{CBS} + Ae^{-\alpha\sqrt{n}}, \tag{1}$$

with A, α, and the CBS-extrapolated energy E_{HF}^{CBS} being parameters to be determined from a least-squares fitting algorithm that has been applied individually for each conformer, and n denoting the cardinal number of the respective basis set, *i.e.* $n = 2$ for ZORA-def2-SVP, $n = 3$ for ZORA-def2-TZVPP, and $n = 4$ for ZORA-def2-QZVPP. A similar extrapolation scheme has also been laid out for the correlation energies, this time following the form proposed by Truhlar:[62]

$$E_{corr}^n = E_{corr}^{CBS} + Bn^{-\beta}, \tag{2}$$

again with B, β, and the CBS-extrapolated energy E_{corr}^{CBS} being parameters to be determined from a least-squares fitting algorithm as before. Extrapolation using all three basis set families (*i.e.* $n = 2, 3, 4$) has been found to yield inconsistent results between the different systems depicted in Figure 1. Hence, extrapolation was laid out using only ZORA-def2-TZVPP and ZORA-def2-QZVPP, *i.e.* an effective two-point extrapolation scheme assuming $\beta = 3$, as originally proposed by Halkier *et al.*[63]

Finally, for systems microsolvated with a Zn^{2+} cation, the slow-converging correlation

contribution of the zinc electrons may lead to large and conformation dependent basis set superposition errors (BSSE).[18,19] To account for that and prior to performing CBS extrapolation as described above, we subject the HF and correlation energies of each Zn^{2+} coordinated conformation to a counterpoise correction as proposed by Boys and Bernardi:[64] Assuming rigid conformers, the BSSE is estimated as

$$
\begin{aligned}
E_{\text{BSSE}} =& E_{\text{BSSE}}(\text{AcH}) + E_{\text{BSSE}}(Zn^{2+}) \text{ , with} \\
& E_{\text{BSSE}}(\text{AcH}) = E^{\text{AcH}+\text{Zn}^{2+}}(\text{AcH}) - E^{\text{AcH}}(\text{AcH}) \text{ , and} \\
& E_{\text{BSSE}}(\text{Zn}^{2+}) = E^{\text{AcH}+\text{Zn}^{2+}}(\text{Zn}^{2+}) - E^{\text{Zn}^{2+}}(\text{Zn}^{2+}),
\end{aligned}
\tag{3}
$$

where $E^{\text{AcH}+\text{Zn}^{2+}}(\text{AcH})$ represents the energy of AcH evaluated in the union of the basis sets on AcH and Zn^{2+}, $E^{\text{AcH}}(\text{AcH})$ represents the energy of AcH evaluated in the basis set of AcH, *etc.* The individual BSSE errors were then subtracted from the HF and correlation energy, respectively.

Empirical force fields

Single-point energy calculations using several out-of-the-box force fields (FFs) were carried out using the `TINKER`[31] molecular modeling package. Two classes of FFs were tackled: (i) conventional FFs, in particular AMBER-99,[34] CHARMM22,[35] and OPLS-AA,[36,37] as well as (ii) polarizable atomic multipole-based FFs that use atomic charge multipole expansion instead of fixed point charges. In particular, these are the 2009 AMOEBA biopolymer FF named AMOEBA-BIO09,[32,33] and the 2013 AMOEBA protein FF named AMOEBA-PRO13.[65] Note that only for systems containing a deprotonated carboxyl group $(-\text{COO}^-)$, parameters were available for all force fields out-of-the-box. As AMOEBA-BIO09 was the only FF available providing also parameters for the neutral carboxyl group $(-\text{COOH})$, FF calculations for systems containing neutral AcH (lower row in Figure 1) were only laid out using this particular FF.

11

Semi-empirical quantum chemistry methods

Semi-empirical quantum chemistry methods are based on the Hartree-Fock method, but follow a simplification strategy by making approximations for computationally demanding terms. In order to account for caused errors, empirical parameters are incorporated into the formalism and fitted against experimental data or high-level calculations.[66] All semi-empirical methods tackled in this work are based on the neglect of diatomic differential overlap (NDDO),[67] a method for approximating computational costly three-center and four-center two-electron integrals. In particular, the different applied models are the Austin Model 1 (AM1),[68] the Parametric Method 3 (PM3),[69] the Parametric Method 6 (PM6),[70] and the Parametric Method 7 (PM7).[71] All semi-empirical method calculations have been carried out using the `MOPAC2016`[72] semi-empirical quantum chemistry program. For the specific case of PM6, two additional long-range dispersion correction schemes were tackled as well. In particular, these are Grimme's D3 correction for dispersion[73] plus a simple function for hydrogen bonds, as well as the corrections to hydrogen bonding and dispersion by Řezáč and Hobza,[74,75] labeled D3H4. The corresponding conjunctive methods are then accordingly being labeled PM6-D3 and PM6-D3H4.

While single-point energy calculations carried out for all other methods in this work refer to total energies on the potential-energy surface, semi-empirical methods yield heats of formation. The heat of formation is defined as the sum of the electronic energy, the nuclear-nuclear repulsion energy, the ionization energy for the valence electrons, the total heat of atomization of all the atoms in the system, and – if available – the energy from hydrogen bonds and dispersion correction.[76] When comparing potential energies of other computational methods with heats of formation obtained from semi-empirical calculations through the means of MAEs and MEs, the systematic shift between the two is accounted for, as explained below.

Density-functional approximations

While density-functional theory (DFT)[77,78] in itself is an exact method, in practice approximations have to be made because the exact form of the exchange-correlation (xc) functional is unknown, except for the free electron gas. A large variety of different DFAs exist, commonly classified into different types depending on the features and formal properties of the xc functionals in question.[79] The ones selected in this work are summarized in the following:

- *Generalized gradient approximations (GGAs)* are characterized by the dependence of the xc functional only on the electron density and its gradient. In this work, we studied the accuracy of the Perdew-Burke-Ernzerhof (PBE)[42] and Becke-Lee-Yang-Parr (BLYP)[15,16] xc functionals.

- In addition to GGAs, *meta-GGAs* also depend on the Laplacian of the electron density or include the kinetic energy density. We tested the M06-L[7] and M11-L[80] xc functionals from the group of Minnesota functionals, as well as the SCAN[81] functional.

- For the computationally more costly class of *hybrid* functionals, the exchange parts of the functional are admixed with exact exchange from Hartree-Fock theory. We tested the PBE0,[47] B3LYP[11,12] and SCAN0[82] functionals. In addition, several hybrid functionals from the group of Minnesota functionals were tested as well, in particular the M06,[83] M06-2X,[83] M08-SO,[84] M08-HX,[84] and M11[85] functionals.

Calculations for the PBE, BLYP, M11-L, SCAN, PBE0, B3LYP, M08-SO, M08-HX, and M11 xc functionals were carried out with `FHI-aims` using `tier 2` basis sets and `really_tight` settings, and including a relativistic treatment by applying the atomic ZORA method.[38] The SCAN and SCAN0 functionals are implemented in `FHI-aims` via the `dfauto` program.[86] Calculations for the M06-L, M06, and M06-2X xc functionals were carried out with `ORCA`, including ZORA and the relativistically recontracted ZORA-def2-QZVPP basis set, as explained above.

Commonly applied semi-local DFAs and conventional hybrid functionals are unable to capture the essence of long-range dispersion effects.[87] Many systems containing biomolecules rely on *van der Waals* interaction treatments for an accurate energetic description.[23,24,88,89] To that end, computationally cheap correction schemes exist that are evaluated *a posteriori*, *i.e.* they are accounted for after the electron density has been obtained via the self-consistent treatment in DFT. Three different vdW correction schemes were tackled in this work:

- The general empirical pairwise additive D3 dispersion correction method by Grimme *et al.*[73] provides a consistent description across the whole periodic table. Here we used the zero-damping function for short ranges, including three-body dispersion contributions. In order to match the long- and midrange correlation of D3 with the semilocal correlation computed by the xc functional, the parameterization of the damping function depends on the xc functional itself. Hence, only xc functionals where an out-of-the-box D3 treatment was available, were tested. In particular, we evaluated M06-L+D3, M06+D3, and M06-2X+D3 using `ORCA`, using the same settings as described above. For the methods of PBE+D3, BLYP+D3, PBE0+D3, and B3LYP+D3, long-range dispersion calculations were done on top of the `FHI-aims` calculated energies using Grimme's stand-alone program `DFT-D3`.[90]

- The parameter-free pairwise Tkatchenko-Scheffler van der Waals scheme (vdW)[43] relies on summing interatomic pairwise, electron-density derived C_6 coefficients, and accurate reference data for the free atoms. As the method is implemented in `FHI-aims`, calculations are carried out for the methods of PBE+vdW, BLYP+vdW, PBE0+vdW, and B3LYP+vdW.

- In contrast to the previous pairwise Tkatchenko-Scheffler scheme that ignores the intrinsic many-body nature of correlation effects, the many-body dispersion scheme labeled MBD[48] (and sometimes also labeled MBD* or MBD@rsSCS) combines the TS scheme with the self-consistent screening (SCS) equation of classical electrodynamics.

In addition, a range-separation (rs) technique is applied, separating correlation into a short-range and a long-range contribution. Calculations were carried out for the methods of PBE+MBD and PBE0+MBD using `FHI-aims`.

In order to avoid high computational costs of hybrid xc functionals and still yield accurate results, recent focus has been set on "low-cost" DFT based *composite electronic structure approaches*. In particular, the PBEh-3c method by Grimme *et al.*[91] aims to efficiently compute structures and interaction energies. It is based on a modified hybrid variant of the PBE GGA xc functional with a relatively large amount of non-local Hartree-Fock-exchange (42%). The orbitals are expanded in Ahlrichs-type valence-double zeta atomic orbital Gaussian basis sets. Furthermore, the D3 scheme and the global counterpoise-correction scheme (gCP)[92] are applied in order to account for long-range dispersion and BSSE effects, respectively. Calculations were carried out with `ORCA`.

Finally, *double hybrid xc functionals* extend hybrid xc functionals in a way that both the exchange and the correlation part contain non-local orbital-dependent components. In particular, we test the B3LYP+XYG3[93] method implemented in `FHI-aims`. As the name implies, the exchange parts of the functional are admixed with exact exchange from Hartree-Fock theory, while a fraction of the correlation part is calculated using Görling-Levy coupling-constant perturbation expansion to the second order (PT2).[94] Calculations were carried out with `FHI-aims` using numerically tabulated atom-centered orbital triple-zeta basis sets with valence-correlation consistency, labeled NAO-VCC-nZ.[95] Constructed analogous to Dunning's correlation-consistent polarized valence-only basis sets (cc-pVnZ),[96] these basis sets utilize the more flexible shape of NAOs. Zhang *et al.* showed that XYG3 provides best results in combination with the triple-zeta NAO-VCC-3Z basis set.[97] Because the NAO-VCC-3Z basis set was not available out-of-the-box for the element of Zn, we used Dunning's analogous cc-pV3Z[98] basis set instead for this particular element.

Mean absolute error and maximum error

In order to compare the energetic performance of different methods, single-point energy calculations of a set of different conformers were compared by means of mean absolute errors (MAEs) and maximum errors (MEs). MAEs of relative energies between the reference method and the method to be benchmarked were calculated as follows:

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |\Delta E_i^{\text{reference}} - \Delta E_i^{\text{benchmarked}} + c|, \tag{4}$$

where the index i runs over all N conformations of a given data set. ΔE_i in principle denotes the energy difference between conformer i and the lowest-energy conformer of the set. The adjustable parameter c is used to systematically shift the reference and benchmark conformational hierarchies versus one another to obtain the lowest possible MAE, rendering the reported MAE value independent of the choice of any reference structure. Similarly, MEs were calculated as follows:

$$ME = \max_{i \in N} |\Delta E_i^{\text{reference}} - \Delta E_i^{\text{benchmarked}} + c|, \tag{5}$$

using the same notation as above. Figure 2 shows an example of a correlation plot including a graphical illustration of $|\Delta E_i^{\text{reference}} - \Delta E_i^{\text{benchmarked}}|$.

Results

Energy hierarchies

Figure 3 shows the obtained energy hierarchies at the PBE0+MBD level after having completed the conformational search for each individual protonation state of bare negatively charged AcH and bare neutral AcH, as well as both systems in presence of a Zn^{2+} cation.

Comparing the two protonation states for negatively charged AcH, $i.e.$ $AcH(N_{\delta 1})-COO^-$
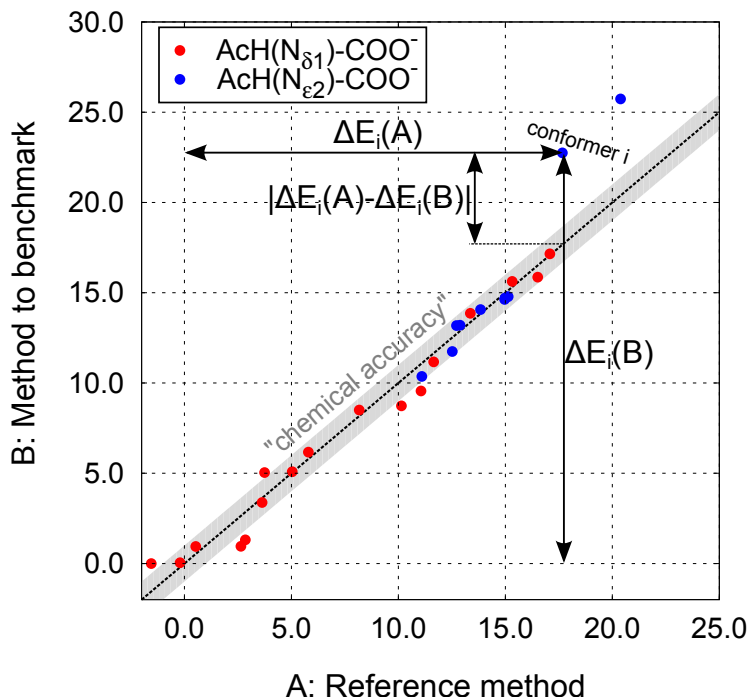
Figure 2: Example of a correlation plot of two different sets of conformers (red and blue). The reference (A) and benchmark (B) conformational hierarchies have already been shifted to minimize the MAE. If the energy description between the reference method and the method to be evaluated agreed perfectly, all points would align on the dashed diagonal line. The gray shading denotes an absolute energy deviation of $1\,\mathrm{kcal/mol}$, *i.e.* the region of "chemical accuracy". For a specific conformer i, the absolute energy deviation $|\Delta E_i^{\mathrm{reference}} - \Delta E_i^{\mathrm{benchmarked}}| = |\Delta E_i(A) - \Delta E_i(B)|$ is illustrated.
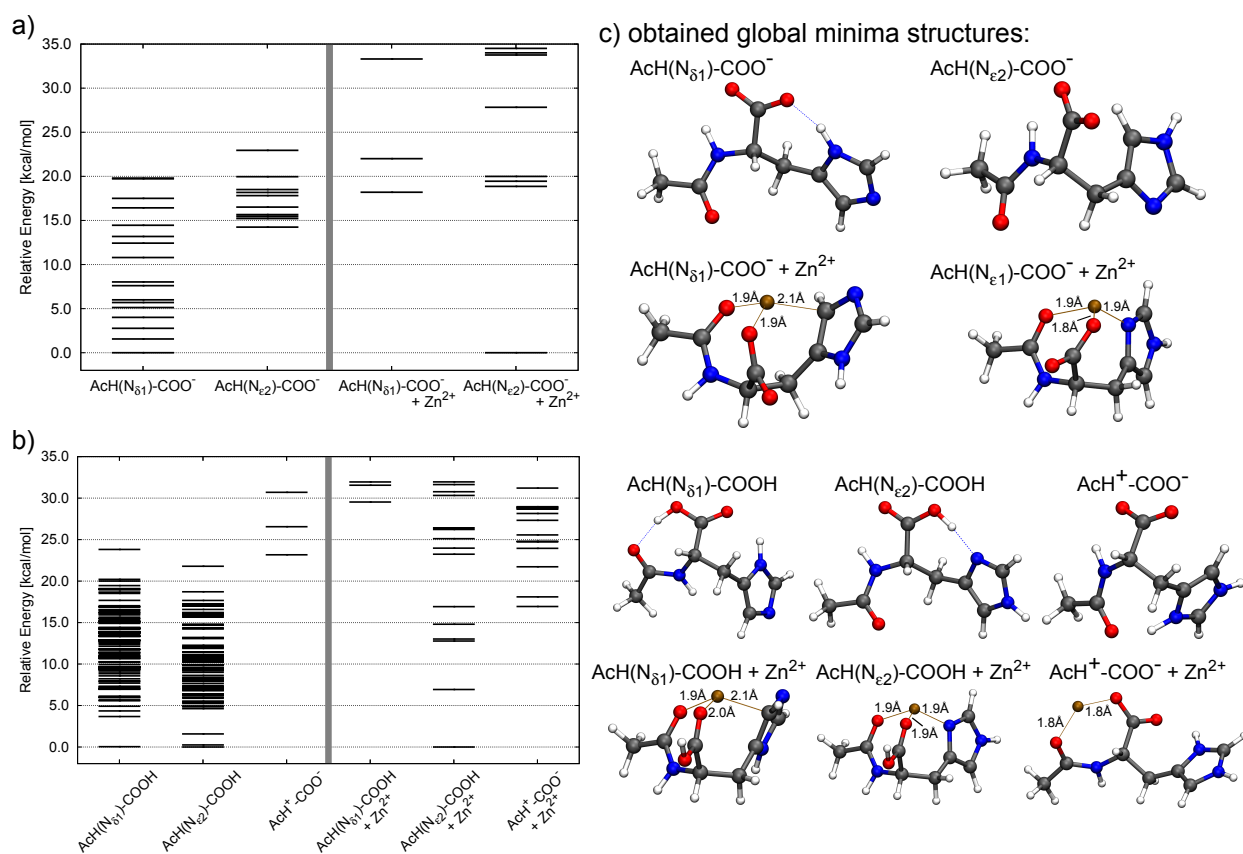
Figure 3: Obtained energy hierarchies at the PBE0+MBD level after having completed the conformational search for (a) negatively charged AcH, bare and with an additional Zn^{2+}, and (b) neutral AcH, bare and with an additional Zn^{2+}. (c) For each depicted protonation state, the geometrical structure of the lowest-energy conformer is illustrated.

and $AcH(N_{\epsilon2})-COO^-$ (see Figure 3(a)), it is immediately evident that the protonation of the nitrogen atoms of the imidazole ring has a large impact concerning energy and structure of the system. The lowest-energy conformer for $AcH(N_{\delta1})-COO^-$ lies 14.3 kcal/mol lower in energy than the lowest-energy conformer for $AcH(N_{\epsilon2})-COO^-$, meaning that the tautomeric state of having the $N_{\delta1}$ nitrogen atom of the imodazole ring protonated is energetically much more favorable that having the $N_{\epsilon2}$ nitrogen atom protonated. The reason for that comes abundantly clear when comparing the two lowest-energy conformers: In the case of $AcH(N_{\delta1})-COO^-$, there exists the geometrical possibility of forming a hydrogen bond between one oxygen of the carboxylate anion at the C-terminus and the nitrogen-bound hydrogen. In case of having the $N_{\epsilon2}$ nitrogen atom protonated, a hydrogen bond cannot be formed as the proton "points away" from the carboxylate anion, explaining the much higher energy of this structure in comparison with its tautomeric counterpart.

The situation however changes drastically when introducing a Zn^{2+} cation to the system. As seen in Figure 3(a), the lowest-energy conformer of $AcH(N_{\delta1})-COO^- + Zn^{2+}$ is now 18.2 kcal/mol higher in energy than the lowest-energy conformer of $AcH(N_{\epsilon2})-COO^- + Zn^{2+}$. The structures look fairly similar in part as the oxygen atom of the carbonyl group at the acetylated N-terminus as well as one oxygen of the carboxylate anion are coordinated towards the Zn^{2+}. They differ however in the different orientation of the imidazole ring towards the cation. In the case of $AcH(N_{\epsilon2})-COO^- + Zn^{2+}$, the deprotonated $N_{\delta1}$ atom allows for a coordinate bonding interaction with the Zn^{2+} cation, resulting in an energetically more favorable structure compared to $AcH(N_{\delta1})-COO^- + Zn^{2+}$ where the deprotonated $N_{\epsilon2}$ atom points away from the cation, resulting in an energetically less favorable cation-π interaction between imidazole ring and cation. Adding a Zn^{2+} cation to the system also results in an increased energetic gap between conformers. For example, the two lowest-energy conformers of $AcH(N_{\delta1})-COO^-$ are separated by 1.6 kcal/mol while the gap increases to 3.8 kcal/mol for $AcH(N_{\delta1})-COO^- + Zn^{2+}$. For $AcH(N_{\epsilon2})-COO^-$, the two lowest-energy conformers are separated by 1.0 kcal/mol, while the gap increases to 18.9 kcal/mol for $AcH(N_{\delta1})-COO^- +$

Zn^{2+}.

The hierarchies of the three different protonation states of bare neutral AcH are shown in Figure 3(b). The global-minimum conformers of the systems of $\text{AcH}(\text{N}_{\delta 1})-\text{COOH}$ and $\text{AcH}(\text{N}_{\epsilon 2})-\text{COOH}$ are very similar in energy, differing only by $0.04\,\text{kcal/mol}$. For $\text{AcH}(\text{N}_{\epsilon 2})-\text{COOH}$, a hydrogen bond is possible between the deprotonated $\text{N}_{\delta 1}$ atom and said proton, resulting in a very similar structure compared to system $\text{AcH}(\text{N}_{\delta 1})-\text{COO}^-$. For $\text{AcH}(\text{N}_{\delta 1})-\text{COOH}$, due to the protonated $\text{N}_{\delta 1}$ atom, the proton at the carboxyl group points away from the imidazole ring and is coordinated towards the N-terminus, forming a hydrogen bond with the carbonyl group. A protonated imidazole ring, as seen in the protonation state of system $\text{AcH}^+-\text{COO}^-$, results in an energetically unfavorable structure, being $23.2\,\text{kcal/mol}$ higher in energy than the global minimum of system $\text{AcH}(\text{N}_{\epsilon 2})-\text{COOH}$.

The situation changes again when introducing a Zn^{2+} cation to the system, see Figure 3(b). The system of $\text{AcH}(\text{N}_{\epsilon 2})-\text{COOH}+\text{Zn}^{2+}$ is energetically most favorable as the structure of the global minimum is very similar to the one of the system of $\text{AcH}(\text{N}_{\epsilon 2})-\text{COO}^- + \text{Zn}^{2+}$: The deprotonated $\text{N}_{\delta 1}$ atom allows for a coordinate bonding interaction with the Zn^{2+} cation that in turn is also coordinated towards the electronegative oxygen atoms at the carboxyl group at the C-terminus and the carbonyl group at the N-terminus. The global minimum of $\text{AcH}^+-\text{COO}^- +\text{Zn}^{2+}$ is $16.9\,\text{kcal/mol}$ higher in energy than the global minimum of $\text{AcH}(\text{N}_{\epsilon 2})-\text{COOH} + \text{Zn}^{2+}$. The positively charged cation and the protonated imidazole ring share no proximity, resulting in a lowest-energy structure where the Zn^{2-} is coordinated between the oxygen of the carbonyl group at the N-terminus and one oxygen of the carboxyl group at the C-terminus. The structure of the global minimum for $\text{AcH}(\text{N}_{\delta 1})-\text{COO}^- +\text{Zn}^{2+}$ is very similar to the one for $\text{AcH}(\text{N}_{\epsilon 2})-\text{COO}^- + \text{Zn}^{2+}$, safe the twisted imidazole ring due to the protonated $\text{N}_{\delta 1}$ atom. Similarly to system $\text{AcH}(\text{N}_{\delta 1})-\text{COO}^- + \text{Zn}^{2+}$, this results in an energetically less favorable cation-π interaction between imidazole ring and cation as the global minimum is $29.5\,\text{kcal/mol}$ higher in energy than the global minimum of $\text{AcH}(\text{N}_{\epsilon 2})-\text{COO}^- + \text{Zn}^{2+}$.

Selection of minima structures

For every protonation state, we selected the lowest-energy structures from the previous global minimum search based on energy criteria. For one, this ensures an emphasis on the most likely structures also seen in experiment as there will always be a bias towards structures with low energy, ignoring individual set-ups or experimental conditions. However, benchmark calculations were done including all possible protonation states for a given overall system charge, except for the case of FFs as explained above. The large energetic differences between global minima (and consequently other low-energy conformers) of individual protonation states, as seen in Figure 3, therefore provides a challenging benchmark testing situation for the different methods. Table 1 summarizes the different energy selection criteria across the systems and protonation states tackled in this work.

Table 1: Minima selection criteria

System	Net charge	Energy cut-off	# Minima
$AcH(N_{\delta 1})-COO^-$	-1	$23.0\,kcal/mol$	17^*
$AcH(N_{\epsilon 2})-COO^-$			10^*
$AcH(N_{\delta 1})-COO^- + Zn^{2+}$	$+1$	$41.5\,kcal/mol$	9
$AcH(N_{\epsilon 2})-COO^- + Zn^{2+}$			9
$AcH(N_{\delta 1})-COOH$	0	$7.0\,kcal/mol$	11
$AcH(N_{\epsilon 2})-COOH$			18
AcH^+-COO^-		$50.0\,kcal/mol$	8^*
$AcH(N_{\delta 1})-COOH + Zn^{2+}$	$+2$	$46.0\,kcal/mol$	9
$AcH(N_{\epsilon 2})-COOH + Zn^{2+}$			18
$AcH^+-COO^- + Zn^{2+}$			22

The "energy cut-off" refers to the energy relative to the respective global minimum for a given total system charge (*i.e.* taking into account all possible protonation states, compare Figure 3), within which all found minima are taken into account. The last column denotes the number of minima used for benchmarking. Numbers denoted with an asterisk (*) mean all found minima for this particular protonation state are considered.

Validation of DLPNO-CCSD(T) as the reference method

As described above and in order to validate DLPNO-CCSD(T) as the reference method used in this work, we checked the consistency of the method against conventional CCSD(T), com-
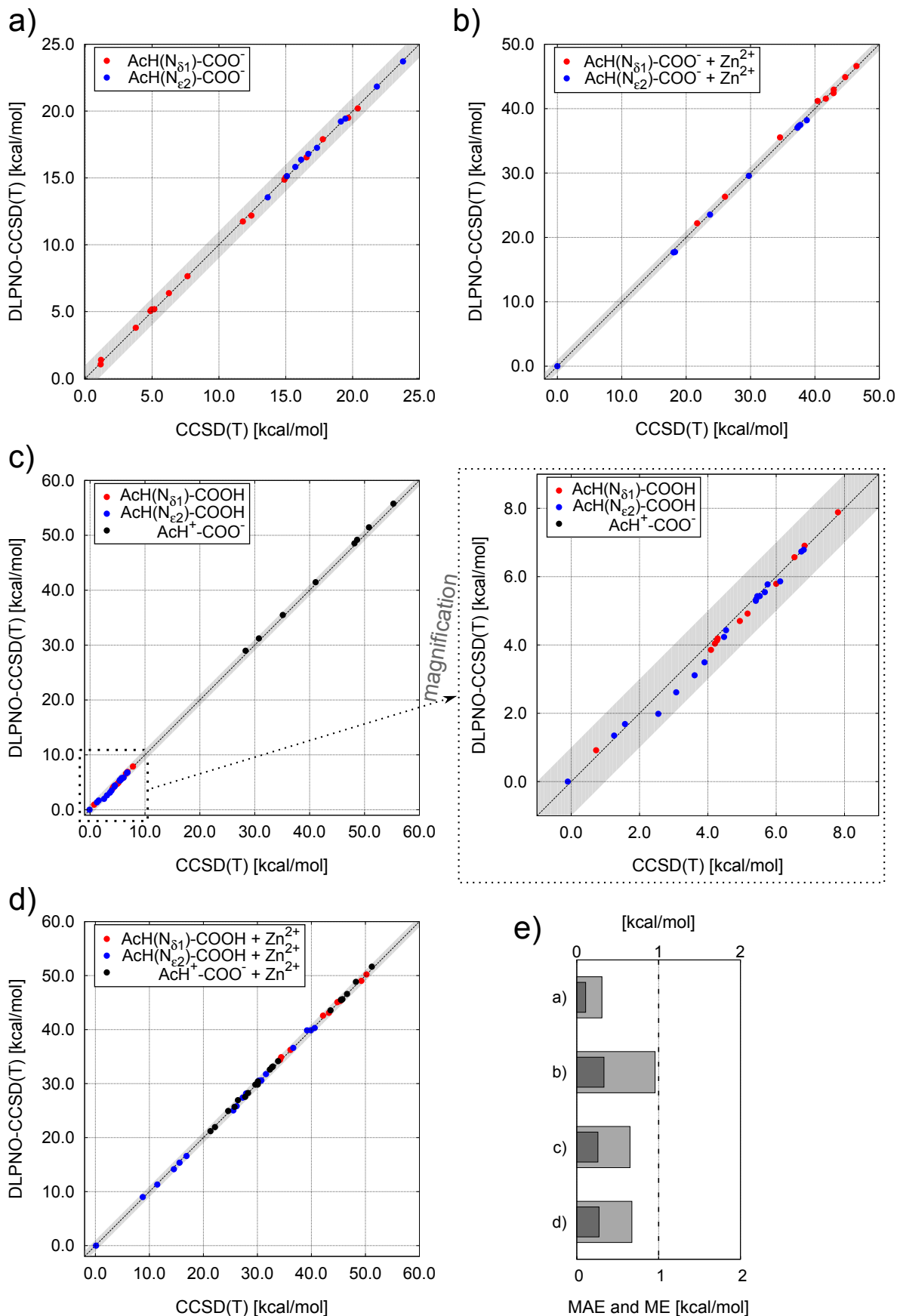
Figure 4: Correlation plots for benchmarking DLPNO-CCSD(T) against conventional CCSD(T) using the ZORA-def2-SVP basis set. The systems tackled refer to (a) negatively charged AcH, (b) the same protonation states in presence of a Zn^{2+} cation, (c) bare neutral AcH, and (d) the same protonation states in presence of a Zn^{2+} cation. The gray

monly referred to as the "golden standard of quantum chemistry". Calculations were laid out using Ahlrichs' relativistically recontracted ZORA-def2-SVP basis set for which CCSD(T) calculations are still affordable with respect to computational costs. Consequently, no extrapolation or counterpoise correction had been applied here, as the intent is to compare the "pure" total energetic performances of both methods, which – if similar – will justify applying DLPNO-CCSD(T) "instead of" conventional CCSD(T) using larger basis sets to benchmark the other computational methods. Figures 4(a)-(d) show the corresponding correlation plots for all systems tackled in this work. The alignment of the points near the dashed diagonal line indicates a very similar energy description between the two methods across all systems and protonation states. To quantify that, MAEs and MEs have been computed according to Equations (4) and (5), respectively. For the four different systems, MAEs and MEs are given in Figure 4(e). In all cases, MAEs are well within "chemical accuracy", *i.e.* smaller than $0.5\,$kcal/mol. Furthermore, MEs are also smaller than $1\,$kcal/mol for all systems. Taking into consideration that different protonation states and minima that differ in energy by up to more than $50\,$kcal/mol have been used, we conclude that DLPNO-CCSD(T) serves as a valid reference method for the benchmarking process of other computational methods.

In order to finally yield accurate total energies serving as benchmarks, counterpoise correction has been applied following Equation (3) and extrapolation to the complete basis set limit has been done following Equations (1) and (2) using Ahlrichs' relativistically recontracted ZORA-def2-SVP, ZORA-def2-TZVPP, and ZORA-def2-QZVPP basis sets.

Benchmarking force fields and semi-empirical methods

Figure 5 shows obtained MAEs and MEs calculated according to Equations (4) and (5) for all systems tackled in this work. As explained above, FF performance evaluation has been treated individually for different protonation states.

Considering bare neutral AcH, see Figure 5(a), conventional FFs that make use of fixed point charges are comparable in performance: For $AcH(N_{\delta 1})-COO^-$, MAEs for AMBER-

a) AcH-COO⁻ (neutral imidazole)

bare   **with Zn²⁺**
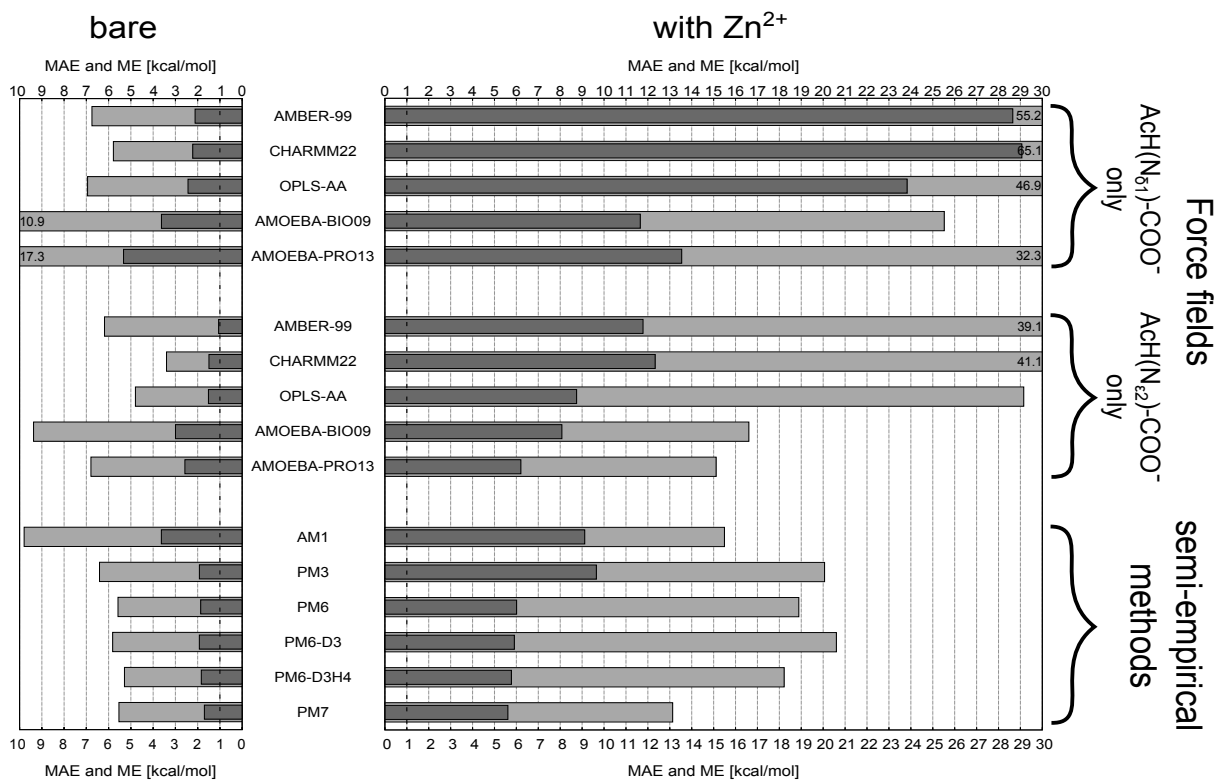
b) AcH-COOH (neutral imidazole) / AcH⁺-COO⁻
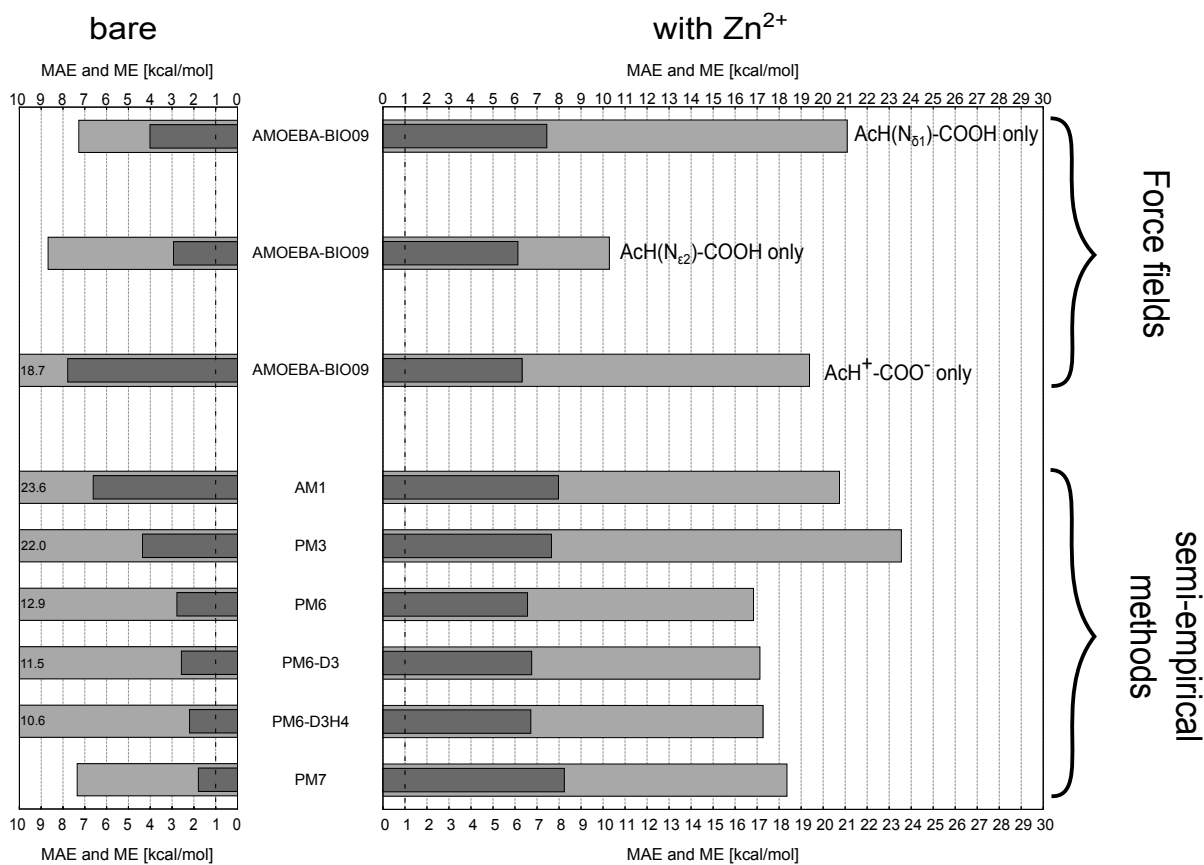
bare   **with Zn²⁺**

24

Figure 5: MAEs (dark-gray) and MEs (light-gray) following Equations (4) and (5) for different force fields and semi-empirical methods with respect to DLPNO-CCSD(T), for which counterpoise correction has been done following Equation (3) and extrapolation to

99, CHARMM22, and OPLS-AA have been found to be 2.1 kcal/mol, 2.2 kcal/mol, and 2.4 kcal/mol, respectively. Considering the fact that FF parameters have been derived from systems in solvation instead of gas-phase calculations applied here, the result can be considered satisfactory. However, large MEs with up to 6.9 kcal/mol for OPLS-AA, indicate a possible large deviation in the energetic description for individual conformers. Somehow surprisingly, polarizable atomic multipole-based FFs AMOEBA-BIO09 and AMOEBA-PRO13 perform worse than their FF counterparts using fixed point charges. Large MEs up to 10.9 kcal/mol and 17.3 kcal/mol for AMOEBA-BIO09 and AMOEBA-PRO13, respectively, indicate severe discrepancies in the energetic description for individual conformers. Consequently, the corresponding MAEs of 3.6 kcal/mol and 5.3 kcal/mol are larger than for conventional FFs.

Qualitative similar results are found for $AcH(N_{\epsilon 2})-COO^-$. Best performance for FFs is found using CHARMM22 with a MAE of 1.5 kcal/mol and a ME of 3.3 kcal/mol.

Semi-empirical methods show a comparable performance to FFs, but are able to describe both protonation states simultaneously, per definition. Best performance is found for PM7 with a MAE of 1.7 kcal/mol and a ME of 5.5 kcal/mol. For PM6, adding a long-range dispersion treatment method, $i.e.$ D3 or D3H4, yields very similar results of approximately 1.9 kcal/mol, as is expected for a system of such small size.

Solvating the system with a single Zn^{2+} cation results in very poor performances for both FFs and semi-empirical methods. Out of the conventional FFs, OPLS-AA shows the best performance with a still very large MAE of 23.8 kcal/mol for $AcH(N_{\delta 1})-COO^- + Zn^{2+}$ and 8.7 kcal/mol for $AcH(N_{\epsilon 2})-COO^- + Zn^{2+}$. Polarizable atomic multipole-based FFs perform slightly better with a MAE of 11.7 kcal/mol using AMOEBA-BIO09 for $AcH(N_{\delta 1})-COO^- + Zn^{2+}$ and a MAE of 6.2 kcal/mol using AMOEBA-PRO13 for $AcH(N_{\epsilon 2})-COO^- + Zn^{2+}$. Semi-empirical methods show a further improvement, with PM7 yielding a MAE of 5.6 kcal/mol and a ME of 13.1 kcal/mol.

As AMOEBA-BIO09 was the only FF available providing parameters out-of-the-box for

the neutral carboxyl group ($-$COOH), FF calculations for systems containing neutral AcH were only laid out using this particular FF, see Figure 5(b). Protonation states with a neutral imidazole ring yield a MAE of 4.0 kcal/mol for AcH($N_{\delta 1}$)$-$COOH and 2.9 kcal/mol for AcH($N_{\epsilon 2}$)$-$COOH. For AcH$^+$$-COO^-$, performance is again very poor yielding a MAE of 7.8 kcal/mol and a ME of 18.4 kcal/mol. When adding a Zn^{2+} cation to the system, the MAE for AMOEBA-BIO09 is larger than 6 kcal/mol for all three protonation states. Out of the semi-empirical methods, PM6 performs best with a MAE of 6.6 kcal/mol.

Benchmarking standard DFAs and methods beyond

Similarly to the previous section, the benchmarking process has been repeated for different kinds of DFAs as well as the wavefunction-based MP2 method. Figure 6 shows obtained MAEs and MEs calculated according to Equations (4) and (5) for all systems tackled in this work. Considering bare neutral AcH, see Figure 6(a), it is interesting to note that all tested methods already provide a very good accuracy as the MAE is less than 1 kcal/mol in all cases. Out of the applied GGA xc functionals, BLYP+D3 shows best performance with a MAE of 0.4 kcal/mol and a ME of 1.1 kcal/mol. It is interesting to see that the applied long-range dispersion schemes all show significant improvement over the methods without such treatment already for systems of such a small size, compare *e.g.* the ME of 3.1 kcal/mol for BLYP with the obtained ME of 1.1 kcal/mol for BLYP+D3. All three different van der Waals treatment methods show a similar performance as the respective obtained MAEs differ by less than 0.1 kcal/mol. Out of the meta-GGA xc functionals, SCAN performs best with a MAE of 0.3 kcal/mol and a ME of 1.0 kcal/mol. Performance of the composite method PBEh-3c is comparable to the bare hybrid xc functional PBE0 with a MAE of 0.8 kcal/mol and a ME of 2.2 kcal/mol. Again, long-range dispersion treatments applied *a posteriori* to the hybrid xc functional calculations improve the performance significantly, compare *e.g.* the ME of 2.7 kcal/mol for B3LYP with the obtained ME of 0.8 kcal/mol for B3LYP+D3. The double hybrid xc functional B3LYP+XYG3 and the wavefunction-based MP2 method

perform equally well with a ME of 0.8 kcal/mol and 0.9 kcal/mol, respectively.
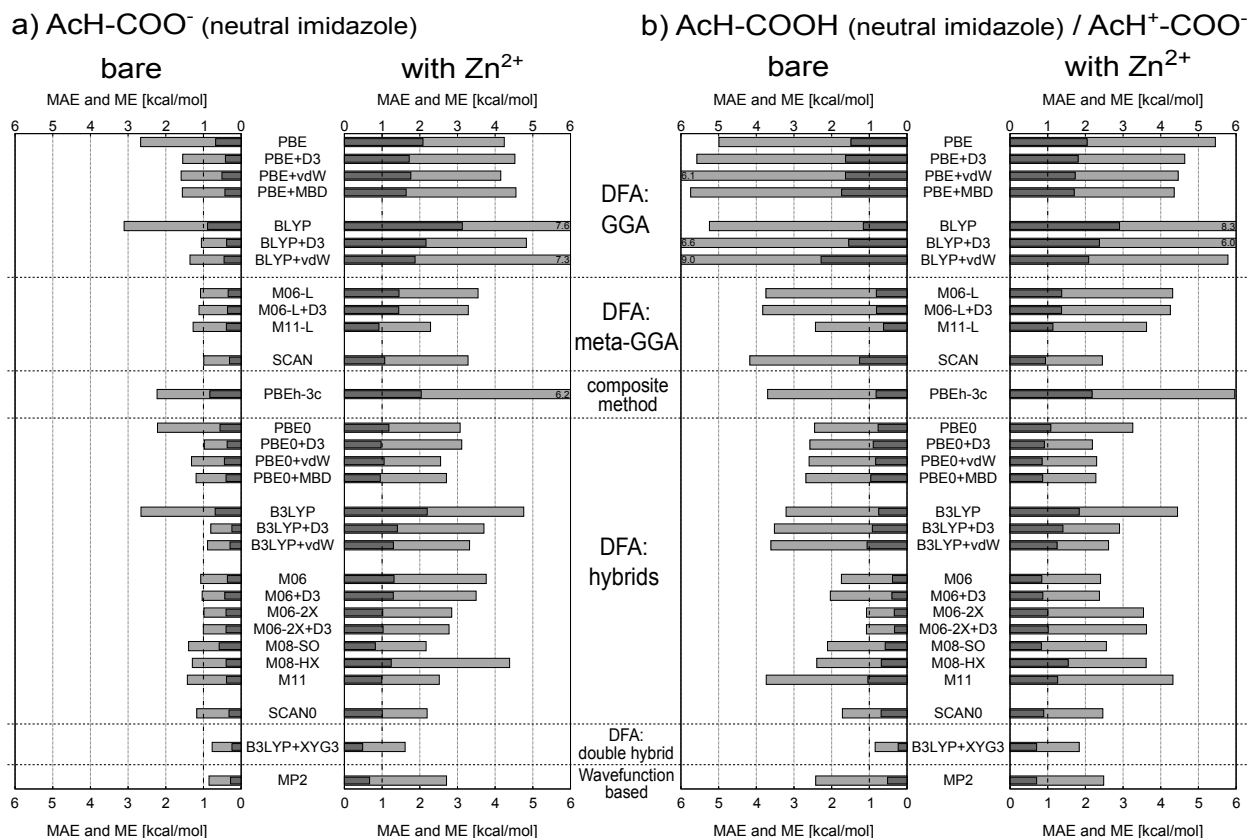

Figure 6: MAEs (dark-gray) and MEs (light-gray) following Equations (4) and (5) for different standard DFAs, the composite method PBEh-3c, double hybrid DFA B3LYP+XYG3, and the wavefunction-based MP2 method with respect to DLPNO-CCSD(T), for which counterpoise correction has been done following Equation (3) and extrapolation to the complete basis set limit has been done following Equations (1) and (2). The tackled systems are (a) negatively charged AcH with and without a Zn^{2+} cation, and (b) neutral AcH with and without a Zn^{2+}.

When adding a Zn^{2+} cation to the system, GGA xc functionals are no longer able to describe the energies within "chemical accuracy". Best performance is found for PBE+MBD with a MAE of 1.6 kcal/mol and a ME of 4.6 kcal/mol. Meta-GGA xc functionals already yield a big improvement as the M11-L xc functional yields a MAE of 0.9 kcal/mol. The composite method PBEh-3c is not sufficient to describe energies of such systems accurately enough as the MAE is found to be 2.0 kcal/mol and a rather large ME of 6.2 kcal/mol is obtained. Hybrid xc functionals provide a generally more accurate energetic description as

PBE0+D3, PBE0+MBD, M06-2X, M06-2X+D3, M08-SO, M11 , and SCAN0 yield MAEs within 1.0 kcal/mol. MP2 yields a MAE of 0.7 kcal/mol. Out of all methods, the double hybrid xc functional B3LYP+XYG3 performs best with a MAE of 0.5 kcal/mol and a ME of 1.6 kcal/mol.

For neutral bare AcH, see Figure 6(b), the benchmarking process is much more challenging as three different protonation states are considered, as well as minima that differ in energy by up to more than 50 kcal/mol. Hence, GGA xc functionals are not able to yield MAEs within "chemical accuracy". Best performance is seen for BLYP with a MAE of 1.2 kcal/mol and a ME of 5.2 kcal/mol. Meta-GGA xc functionals already show a big improvement with M11-L giving the best performance with a MAE of 0.6 kcal/mol. The composite method PBEh-3c also yields a small MAE of 0.8 kcal/mol while the corresponding ME of 3.7 kcal/mol indicates that larger energetic deviations are possible for individual conformers. Hybrid xc functionals again perform very well as all MAEs are within 1.0 kcal/mol. Best performance is found for M06-2X with a MAE of 0.3 kcal/mol and a ME of 1.1 kcal/mol. Out of all methods, best performance is again found for B3LYP+XYG3 with a MAE of 0.2 kcal/mol and a ME of 0.8 kcal/mol.

When adding a Zn^{2+} cation to the system, performance of the methods is comparable to $AcH^+-COO^- + Zn^{2+}$. GGA xc functionals all yield a MAE above 1 kcal/mol. In order to reach "chemical accuracy" one needs to rely on meta-GGA where the SCAN functional yields a MAE of 0.9 kcal/mol. Out of the hybrid xc functionals, PBE0+D3, PBE0+vdW, PBE+MBD, M06, M06+D3, M06-2X, M06-2X+D3, M06-SO, and SCAN0 yield MAEs within 1.0 kcal/mol. Out of all methods, best performance is again found for B3LYP+XYG3 with a MAE of 0.7 kcal/mol and a ME of 1.8 kcal/mol.

Considering calculation times

For applications, one not only needs to consider the accuracy of a particular method, but also the required computational costs and times. All FF and semi-empirical calculations

in this work have been laid out on a single CPU core and took between $0.1\,s$ and $0.3\,s$ per single-point energy evaluation. Timings of these methods are all similar due to the small size of the benchmark systems. Because of the fast timings of energy evaluations, conventional FFs are applied if an excessive amount of single-point energy evaluations is required, *e.g.* for molecular dynamics simulations or conformational searches.[99] However, as seen in this work where this energy description model was found only acceptable for bare neutral AcH, one should generally cross-check with other more accurate methods.

Concerning DFT calculations, timings depend on the applied xc functional, used basis set, the system, and the implementation of the method itself. On a machine with 32 CPU cores and for the system of $AcH(N_{\delta 1})-COOH + Zn^{2+}$, it took $43\,s$ on average for a single-point energy calculation with `FHI-aims` applying the GGA xc functionals PBE and BLYP, using `tier 2` basis sets and `really_tight` settings. Using the SCAN xc functional and the M11-L meta-GGA xc functional with the same settings took $69\,s$ and $97\,s$ on average, respectively. Calculations for the two best performant hybrid xc functionals M08-SO and SCAN0 took $839\,s$ and $586\,s$ on average using the same settings.

However, for most DFT production purposes one would not rely on computationally costly, yet very accurate, `really_tight` settings, as done in this work. For standard cases, `tight` settings in combination with `tier 2` basis sets already provide meV-level accurate energy differences,[38] *i.e.* within $0.02\,kcal/mol$. Indeed, repeating the procedure for the PBE, BLYP, PBE0, and B3LYP xc functionals but using `tight` settings yields virtually identical results. On a machine with 32 CPU cores and for the system of $AcH(N_{\delta 1})-COOH + Zn^{2+}$, computational time gets then reduced from $43\,s$ to $27\,s$ on average for a single-point energy calculation applying the GGA xc functionals PBE. Similarly for the hybrid xc functional PBE0, average calculation times of $738\,s$ with `really_tight` settings get reduced to $727\,s$ with `tight` settings.

The composite method PBEh-3c that gave MAEs within "chemical accuracy" for the systems without a Zn^{2+} cation, took $213\,s$ on average for a single-point energy calculation

on a single CPU core using `ORCA`. The most accurate method across all systems and protonation states, B3LYP+XYG3, took 751 s on average for a single-point energy evaluation on a machine with 32 CPU cores using `FHI-aims`. While MAEs for MP2 are comparable with B3LYP+XYG3 and within "chemical accuracy", energy evaluation times are much larger due to the large basis sets required for accurate predictions. On a machine with 32 CPU cores, it took 2242 s on average for an MP2 energy calculation using `ORCA` and the ZORA-def2-QZVPP basis set.

Conclusions

We examined the goodness of commonly applied levels of theory, *i.e.* force fields, semi-empirical methods, density-functional approximations (DFA), and wavefunction-based methods with respect to high-level coupled-cluster calculations. To that end, benchmark systems consisting of either a bare acetylhistidine or microsolvated with a Zn^{2+} cation were (i) conformationally sampled by performing a global energy minimum search combining both FF and DFA, and (ii) obtained conformational minima were used for benchmarking against DLPNO-CCSD(T) single-point energy-calculations.

For bare negatively charged AcH, the obtained energy hierarchies on the hybrid DFA level showed that the protonation state of $AcH(N_{\delta 1})-COO^-$ is energetically favorable compared to $AcH(N_{\epsilon 2})-COO^-$ as the respective global minima differ by 14.3 kcal/mol in energy. The situation is reversed when adding a Zn^{2+} cation to the system: the protonation state of $AcH(N_{\epsilon 2})-COO^- + Zn^{2+}$ is energetically preferred as the respective global minima differ by 18.2 kcal/mol. Considering bare neutral AcH, the two protonation states of $AcH(N_{\delta 1})-COOH$ and $AcH(N_{\epsilon 2})-COOH$ yield global minima that are similar in energy. When adding a Zn^{2+} cation to the system, $AcH(N_{\epsilon 2})-COO^- + Zn^{2+}$ is energetically preferred to the other two protonation states of $AcH^+-COO^- + Zn^{2+}$ and $AcH(N_{\delta 1})-COO^- + Zn^{2+}$, as the global minima differ by 16.9 kcal/mol and 29.5 kcal/mol in energy, respectively.

The benchmarking process, based on single-point energy calculations and assessed by means of MAEs and MEs, revealed that force fields and semi-empirical methods are generally not reliable enough for an energetic description of these systems within "chemical accuracy" of $1\,\mathrm{kcal/mol}$. While GGA xc functionals like PBE and BLYP, as well as the composite method PBEh-3c have problems in their energetic description for systems containing a Zn^{2+} cation, it is possible to reach "chemical accuracy" for all systems already using the meta-GGA SCAN xc functional. Hybrid xc functionals perform generally well with MAEs within $1\,\mathrm{kcal/mol}$ for most of them. Out of the hybrid xc functionals, best performance is shown for M06-SO and SCAN0. Out of all tested methods, the double hybrid xc functional B3LYP+XYG3 resembles the benchmark method DLPNO-CCSD(T) best with a MAE of $0.7\,\mathrm{kcal/mol}$ and a ME of $1.8\,\mathrm{kcal/mol}$. While MP2 performs similarly as B3LYP+XYG3, computational costs, *i.e.* timings, are increased by a factor of 4 in comparison due to the large basis sets required for accurate results.

Acknowledgement

Supporting Information

SI.zip The *zip*-archive includes a collection of all found conformers (*xyz* file format) and corresponding energies for the obtained hierarchies at the PBE0+MBD level after having completed the conformational search, see Figure 3. In addition, the *zip*-archive also includes listings of all calculated single-point energies for all computational methods and benchmark structures. See here:

`http://w0.rz-berlin.mpg.de/user/baldauf/carsten_pdf/Schneider_HisBench_2018_SI.zip`

References

(1) Permyakov, E. A. *Metalloproteomics*; John Wiley & Sons, 2009.

(2) Bertini, I.; Gray, H. B.; Stiefel, E. I.; Valentine, J. *Biological Inorganic Chemistry*; University Science Books, 2007.

(3) Kepp, K. P. Bioinorganic Chemistry of Alzheimer's Disease. *Chem. Rev.* **2012**, *112*, 5193–5239.

(4) Sandstead, H. Understanding Zinc: Recent Observations and Interpretations. *J. Lab. Clin. Med.* **1994**, *124*, 322–7.

(5) Lindskog, S. Structure and Mechanism of Carbonic Anhydrase. *Pharmacol. Therapeut.* **1997**, *74*, 1–20.

(6) Amin, E. A.; Truhlar, D. G. Zn Coordination Chemistry: Development of Benchmark Suites for Geometries, Dipole Moments, and Bond Dissociation Energies and their Use to Test and Validate Density Functionals and Molecular Orbital Theory. *J. Chem. Theory Comput.* **2008**, *4*, 75–85.

(7) Zhao, Y.; Truhlar, D. G. A New Local Density Functional for Main-group Thermo-chemistry, Transition Metal Bonding, Thermochemical Kinetics, and Noncovalent Interactions. *J. Chem. Phys.* **2006**, *125*, 194101.

(8) Møller, C.; Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.* **1934**, *46*, 618–622.

(9) Head-Gordon, M.; Poplei, J. A.; Frisch, M. J. MP2 Energy Evaluation by Direct Methods. *Chem. Phys. Lett.* **1988**, *153*, 503–506.

(10) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. A Fifth-order Perturbation Comparison of Electron Correlation Theories. *Chem. Phys. Lett.* **1989**, *157*, 479–483.

(11) Vosko, S. H.; Wilk, L.; Nusair, M. Accurate Spin-dependent Electron Liquid Correlation Energies for Local Spin Density Calculations: A Critical Analysis. *Can. J. Phys.* **1980**, *58*, 1200–1211.

(12) Scuseria, G. E.; Staroverov, V. N. Progress in the Development of Exchange-correlation Functionals. 2005.

(13) Rayón, V. M.; Valdés, H.; Díaz, N.; Suárez, D. Monoligand Zn(II) Complexes: Ab Initio Benchmark Calculations and Comparison with Density Functional Theory Methodologies. *J. Chem. Theory Comput.* **2008**, *4*, 243–256.

(14) Weaver, M. N.; Merz, K. M.; Ma, D.; Kim, H. J.; Gagliardi, L. Calculation of Heats of Formation for Zn Complexes: Comparison of Density Functional Theory, Second Order Perturbation Theory, Coupled-Cluster and Complete Active Space Methods. *Journal of Chemical Theory and Computation* **2013**, *9*, 5277–5285.

(15) Becke, A. D. Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior. *Phys. Rev. A* **1988**, *38*, 3098–3100.

(16) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-energy Formula into a Functional of the Electron Density. *Phys. Rev. B* **1988**, *37*, 785–789.

(17) Hobza, P. Calculations on Noncovalent Interactions and Databases of Benchmark Interaction Energies. *Accounts Chem. Res.* **2012**, *45*, 663–672.

(18) Jansen, H. B.; Ros, P. Non-empirical Molecular Orbital Calculations on the Protonation of Carbon Monoxide. *Chem. Phys. Lett.* **1969**, *3*, 140–143.

(19) Liu, B.; McLean, A. D. Accurate Calculation of the Attractive Interaction of Two Ground State Helium Atoms. *J. Chem. Phys.* **1973**, *59*, 4557–4558.

(20) Gutten, O.; Beššeová, I.; Rulíšek, L. Interaction of Metal Ions with Biomolecular Ligands: How Accurate Are Calculated Free Energies Associated with Metal Ion Complexation? *J. Phys. Chem. A* **2011**, *115*, 11394–11402.

(21) Gutten, O.; Rulíšek, L. Predicting the Stability Constants of Metal-Ion Complexes from First Principles. *Inorg. Chem.* **2013**, *52*, 10347–10355.

(22) Navrátil, V.; Klusák, V.; Rulíšek, L. Theoretical Aspects of Hydrolysis of Peptide Bonds by Zinc Metalloenzymes. *Chem. Eur. J.* **2013**, *19*, 16634–16645.

(23) Ropo, M.; Schneider, M.; Baldauf, C.; Blum, V. First-principles Data Set of 45,892 Isolated and Cation-coordinated Conformers of 20 Proteinogenic Amino Acids. *Sci. Data* **2016**, *3*, 160009.

(24) Ropo, M.; Blum, V.; Baldauf, C. Trends for isolated amino acids and dipeptides: Conformation, divalent ion binding, and remarkable similarity of binding to calcium and lead. *Sci. Rep.* **2016**, *6*, 35772.

(25) Dudev, T.; Lim, C. Competition among Metal Ions for Protein Binding Sites: Determinants of Metal Ion Selectivity in Proteins. *Chem. Rev.* **2014**, *114*, 538–556.

(26) Dudev, T.; Lim, C. Modeling Zn2+-Cysteinate Complexes in Proteins. *J. Phys. Chem. B* **2001**, *105*, 10709–10714.

(27) Dudev, T.; Lim, C. Metal-Binding Affinity and Selectivity of Nonstandard Natural Amino Acid Residues from DFT/CDM Calculations. *J. Phys. Chem. B* **2009**, *113*, 11754–11764.

(28) Gutten, O.; Rulíšek, L. How simple is too simple? Computational perspective on importance of second-shell environment for metal-ion selectivity. *Phys. Chem. Chem. Phys.* **2015**, *17*, 14393–14404.

(29) Marianski, M.; Supady, A.; Ingram, T.; Schneider, M.; Baldauf, C. Assessing the Accuracy of Across-the-Scale Methods for Predicting Carbohydrate Conformational Energies for the Examples of Glucose and ?-Maltose. *J. Chem. Theory Comput.* **2016**, *12*, 6157–6168.

(30) Ponder, J. W. and Richards, F. M., An Efficient Newton-like Method for Molecular Mechanics Energy Minimization of Large Molecules. *J. Comput. Chem.* **1987**, *8*, 1016–1024.

(31) Ponder, J. W. TINKER - Software Tools for Molecular Design. *Washington University School of Medicine, Saint Louis, MO* **2013**,

(32) Ren, P.; Ponder, J. W. Polarizable Atomic Multipole Water Model for Molecular Mechanics Simulation. *J. Phys. Chem. B* **2003**, *107*, 5933–5947.

(33) Wu, J. C.; Piquemal, J.-P.; Chaudret, R.; Reinhardt, P.; Ren, P. Polarizable Molecular Dynamics Simulation of Zn(II) in Water Using the AMOEBA Force Field. *J. Chem. Theory Comput.* **2010**, *6*, 2059–2070.

(34) Wang, J.; Cieplak, P.; Kollman, P. A. How well does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules? *J. Comput. Chem.* **2000**, *21*, 1049–1074.

(35) MacKerell, A. D. et al. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(36) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.

(37) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Ac-

curate Quantum Chemical Calculations on Peptides. *J. Phys. Chem. B* **2001**, *105*, 6474–6487.

(38) Blum, V.; Gehrke, R.; Hanke, F.; Havu, P.; Havu, V.; Ren, X.; Reuter, K.; Scheffler, M. Ab initio Molecular Simulations with Numeric Atom-centered Orbitals. *Comp. Phys. Commun.* **2009**, *180*, 2175–2196.

(39) Ren, X.; Rinke, P.; Blum, V.; Wieferink, J.; Tkatchenko, A.; Sanfilippo, A.; Reuter, K.; Scheffler, M. Resolution-of-Identity Approach to Hartree-Fock, Hybrid Density Functionals, RPA, MP2 and GW with Numeric Atom-centered Orbital Basis Functions. *New J. Phys.* **2012**, *14*, 053020.

(40) Auckenthaler, T.; Blum, V.; Bungartz, H.-J.; Huckle, T.; Johanni, R.; Krämer, L.; Lang, B.; Lederer, H.; Willems, P. Parallel Solution of Partial Symmetric Eigenvalue Problems from Electronic Structure Calculations. *Parallel Comput.* **2011**, *37*, 783–794.

(41) Nocedal, J.; Wright, S. J. *Numerical Optimization*, 2nd ed.; Springer, 2006.

(42) Perdew, J.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

(43) Tkatchenko, A.; Scheffler, M. Accurate Molecular Van Der Waals Interactions from Ground-State Electron Density and Free-Atom Reference Data. *Phys. Rev. Lett.* **2009**, *102*, 073005.

(44) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An Open Chemical Toolbox. *J. Cheminform.* **2011**, *3*, 33.

(45) Sokal, R.; Michener, C. A Statistical Method for Evaluating Systematic Relationships. *Univ. Kans. Sci. Bull.* **1958**, *38*, 1409–1438.

(46) SciPy: Open Source Scientific Tools for Python. 2001–; `http://www.scipy.org/`.

(47) Adamo, C.; Barone, V. Toward Reliable Density Functional Methods without Adjustable Parameters: The PBE0 Model. *J. Chem. Phys.* **1999**, *110*, 6158–6170.

(48) Ambrosetti, A.; Reilly, A. M.; DiStasio, R. A.; Tkatchenko, A. Long-range Correlation Energy Calculated from Coupled Atomic Response Functions. *J. Chem. Phys.* **2014**, *140*, 18A508.

(49) Čížek, J. On the Correlation Problem in Atomic and Molecular Systems. Calculation of Wavefunction Components in Ursell-Type Expansion Using Quantum-Field Theoretical Methods. *J. Chem. Phys.* **1966**, *45*, 4256–4266.

(50) Purvis III, G. D.; Bartlett, R. J. A Full Coupled-cluster Singles and Doubles Model: The Inclusion of Disconnected Triples. *J. Chem. Phys.* **1982**, *76*, 1910–1918.

(51) Řezáč, J.; Hobza, P. Describing Noncovalent Interactions beyond the Common Approximations: How Accurate Is the "Gold Standard," CCSD(T) at the Complete Basis Set Limit? *J. Chem. Theory Comput.* **2013**, *9*, 2151–2155.

(52) Riplinger, C.; Neese, F. An Efficient and Near Linear Scaling Pair Natural Orbital Based Local Coupled Cluster Method. *J. Chem. Phys.* **2013**, *138*, 034106.

(53) Riplinger, C.; Sandhoefer, B.; Hansen, A.; Neese, F. Natural Triple Excitations in Local Coupled Cluster Calculations with Pair Natural Orbitals. *J. Chem. Phys.* **2013**, *139*, 134101.

(54) Neese, F. The ORCA Program System. *WIREs Comput. Mol. Sci.* **2012**, *2*, 73–78.

(55) 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.

(56) van Lenthe, E.; Baerends, E. J.; Snijders, J. G. Relativistic Regular Two-component Hamiltonians. *J. Chem. Phys.* **1993**, *99*, 4597–4610.

(57) van Wüllen, C. Molecular Density Functional Calculations in the Regular Relativistic Approximation: Method, Application to Coinage Metal Diatomics, Hydrides, Fluorides and Chlorides, and Comparison with First-order Relativistic Calculations. *J. Chem. Phys.* **1998**, *109*, 392–399.

(58) Pantazis, D. A.; Chen, X.-Y.; Landis, C. R.; Neese, F. All-Electron Scalar Relativistic Basis Sets for Third-Row Transition Metal Atoms. *J. Chem. Theory Comput.* **2008**, *4*, 908–919.

(59) Liakos, D. G.; Sparta, M.; Kesharwani, M. K.; Martin, J. M. L.; Neese, F. Exploring the Accuracy Limits of Local Pair Natural Orbital Coupled-Cluster Theory. *J. Chem. Theory Comput.* **2015**, *11*, 1525–1539.

(60) Kossmann, S.; Neese, F. Comparison of Two Efficient Approximate Hartree-Fock Approaches. *Chem. Phys. Lett.* **2009**, *481*, 240–243.

(61) Karton, A.; Martin, J. M. L. Comment on: "Estimating the Hartree-Fock Limit from Finite Basis Set Calculations" [Jensen F (2005) Theor Chem Acc 113:267]. *Theor. Chem. Acc.* **2006**, *115*, 330–333.

(62) Truhlar, D. Basis-set Extrapolation. *Chem. Phys. Lett.* **1998**, *294*, 45–48.

(63) Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. Basis-set Convergence in Correlated Calculations on Ne, N_2, and H_2O. *Chem. Phys. Lett.* **1998**, *286*, 243–252.

(64) Boys, S.; Bernardi, F. The Calculation of Small Molecular Interactions by the Differences of Separate Total Energies. Some Procedures with Reduced Errors. *Mol. Phys.* **1970**, *19*, 553–566.

(65) Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P. Polarizable Atomic

Multipole-Based AMOEBA Force Field for Proteins. *J. Chem. Theory Comput.* **2013**, *9*, 4046–4063.

(66) Thiel, W. Semiempirical Quantum-chemical Methods. *WIREs Comput. Mol. Sci.* **2014**, *4*, 145–157.

(67) Pople, J. A.; Santry, D. P.; Segal, G. A. Approximate Self-Consistent Molecular Orbital Theory. I. Invariant Procedures. *J. Chem. Phys.* **1965**, *43*, S129–S135.

(68) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. Development and Use of Quantum Mechanical Molecular Models. 76. AM1: A New General Purpose Quantum Mechanical Molecular Model. *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.

(69) Stewart, J. J. P. Optimization of Parameters for Semiempirical Methods I. Method. *J. Comput. Chem.* **1989**, *10*, 209–220.

(70) Stewart, J. J. P. Optimization of Parameters for Semiempirical Methods V: Modification of NDDO Approximations and Application to 70 Elements. *J. Mol. Model.* **2007**, *13*, 1173–1213.

(71) Stewart, J. J. P. Optimization of Parameters for Semiempirical Methods VI: More Modifications to the NDDO Approximations and Re-optimization of Parameters. *J. Mol. Model.* **2013**, *19*, 1–32.

(72) Stewart, J. J. P. MOPAC2016. Stewart Computational Chemistry, `http://OpenMOPAC.net`.

(73) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate Ab initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.

(74) Řezáč, J.; Hobza, P. Advanced Corrections of Hydrogen Bonding and Dispersion for

Semiempirical Quantum Mechanical Methods. *J. Chem. Theory Comput.* **2012**, *8*, 141–151.

(75) Vorlová, B.; Nachtigallová, D.; Jirásková-Vaníčková, J.; Ajani, H.; Jansa, P.; Řezáč, J.; Fanfrlík, J.; Otyepka, M.; Hobza, P.; Konvalinka, J.; Lepšík, M. Malonate-based Inhibitors of Mammalian Serine Racemase: Kinetic Characterization and Structure-based Computational Study. *Eur. J. Med. Chem.* **2015**, *89*, 189–197.

(76) Stewart, J. J. P. MOPAC2016 Manual. `http://openmopac.net/manual/SCF_calc_hof.html`.

(77) Hohenberg, P.; Kohn, W., Inhomogeneous Electron Gas. *Phys. Rev.* **1964**, *136*, B864–B871.

(78) Kohn, W.; Sham, L. J., Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140*, A1133–A1138.

(79) Perdew, J. P.; Schmidt, K. Jacob's Ladder of Density Functional Approximations for the Exchange-correlation Energy. *AIP Conf. Proc.* **2001**, *577*, 1–20.

(80) Peverati, R.; Truhlar, D. G. M11-L: A Local Density Functional That Provides Improved Accuracy for Electronic Structure Calculations in Chemistry and Physics. *J. Phys. Chem. Lett.* **2012**, *3*, 117–124.

(81) Sun, J.; Ruzsinszky, A.; Perdew, J. P. Strongly Constrained and Appropriately Normed Semilocal Density Functional. *Phys. Rev. Lett.* **2015**, *115*, 036402.

(82) Hui, K.; Chai, J.-D. SCAN-based Hybrid and Double-hybrid Density Functionals from Models Without Fitted Parameters. *J. Chem. Phys.* **2016**, *144*, 044114.

(83) Zhao, Y.; Truhlar, D. G. The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Noncovalent Interactions, Excited States, and

Transition Elements: Two New Functionals and Systematic Testing of Four M06-class Functionals and 12 Other Functionals. *Theor. Chem. Acc.* **2008**, *120*, 215–241.

(84) Zhao, Y.; Truhlar, D. G. Exploring the Limit of Accuracy of the Global Hybrid Meta Density Functional for Main-Group Thermochemistry, Kinetics, and Noncovalent Interactions. *J. Chem. Theory Comput.* **2008**, *4*, 1849–1868.

(85) Peverati, R.; Truhlar, D. G. Improving the Accuracy of Hybrid Meta-GGA Density Functionals by Range Separation. *J. Phys. Chem. Lett.* **2011**, *2*, 2810–2817.

(86) Strange, R.; Manby, F.; Knowles, P. Automatic Code Generation in Density Functional Theory. *Comp. Phys. Commun.* **2001**, *136*, 310–318.

(87) Grimme, S. Density Functional Theory with London Dispersion Corrections. *WIREs Comput. Mol. Sci.* **2011**, *1*, 211–228.

(88) Baldauf, C.; Rossi, M. Going clean: structure and dynamics of peptides in the gas phase and paths to solvation. *Journal of Physics: Condensed Matter* **2015**, *27*, 493002.

(89) Schubert, F.; Rossi, M.; Baldauf, C.; Pagel, K.; Warnke, S.; von Helden, G.; Filsinger, F.; Kupser, P.; Meijer, G.; Salwiczek, M.; Koksch, B.; Scheffler, M.; Blum, V. Exploring the Conformational Preferences of 20-residue Peptides in Isolation: Ac-Ala19-Lys + H+ vs. Ac-Lys-Ala19 + H+ and the Current Reach of DFT. *Phys. Chem. Chem. Phys.* **2015**, *17*, 7373–7385.

(90) Grimme, S. DFT-D3. `https://www.chemie.uni-bonn.de/pctc/mulliken-center/software/dft-d3/dft-d3`.

(91) Grimme, S.; Brandenburg, J. G.; Bannwarth, C.; Hansen, A. Consistent Structures and Interactions by Density Functional Theory with Small Atomic Orbital Basis Sets. *J. Chem. Phys.* **2015**, *143*, 054107.

(92) Kruse, H.; Grimme, S. A Geometrical Correction for the Inter- and Intra-molecular Basis Set Superposition Error in Hartree-Fock and Density Functional Theory Calculations for Large Systems. *J. Chem. Phys.* **2012**, *136*, 154101.

(93) Zhang, Y.; Xu, X.; Goddard, W. A. Doubly Hybrid Density Functional for Accurate Descriptions of Nonbond Interactions, Thermochemistry, and Thermochemical Kinetics. *P. Natl. Acad. Sci. USA* **2009**, *106*, 4963–4968.

(94) Görling, A.; Levy, M. Correlation-energy Functional and its High-density Limit Obtained from a Coupling-constant Perturbation Expansion. *Phys. Rev. B* **1993**, *47*, 13105–13113.

(95) Zhang, I. Y.; Ren, X.; Rinke, P.; Blum, V.; Scheffler, M. Numeric Atom-centered-orbital Basis Sets with Valence-correlation Consistency from H to Ar. *New J. Phys.* **2013**, *15*, 123033.

(96) Dunning Jr., T. H. Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron through Neon and Hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007–1023.

(97) Zhang, I. Y.; Luo, Y.; Xu, X. Basis Set Dependence of the Doubly Hybrid XYG3 Functional. *J. Chem. Phys.* **2010**, *133*, 104105.

(98) Balabanov, N. B.; Peterson, K. A. Systematically Convergent Basis Sets for Transition Metals. I. All-electron Correlation Consistent Basis Sets for the 3d Elements Sc-Zn. *J. Chem. Phys.* **2005**, *123*, 064107.

(99) Schneider, M.; Masellis, C.; Rizzo, T.; Baldauf, C. Kinetically Trapped Liquid-State Conformers of a Sodiated Model Peptide Observed in the Gas Phase. *J. Phys. Chem. A* **2017**, *121*, 6838–6844.