THE JOURNAL OF PHYSICAL CHEMISTRY A

Article

pubs.acs.org/JPCA

Validation Challenge of Density-Functional Theory for Peptides—Example of Ac-Phe-Ala$_5$-LysH$^+$

Mariana Rossi,*[,†,‡] Sucismita Chutia,[†] Matthias Scheffler,[†] and Volker Blum*[,†,§]

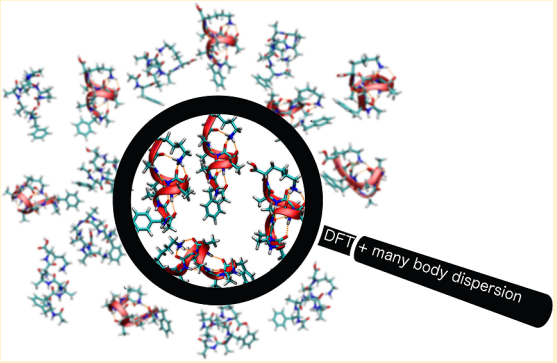
[†]Fritz-Haber-Institut der Max-Planck-Gesellschaft, Berlin, D-14195 Germany

[‡]Physical and Theoretical Chemistry Laboratory, University of Oxford, OX1 3QZ Oxford, U.K.

[§]Department of Mechanical Engineering and Materials Science and Center for Materials Genomics, Duke University, Durham, North Carolina 27708, United States

Supporting Information

ABSTRACT: We assess the performance of a group of exchange-correlation functionals for predicting the secondary structure of peptide chains, up to a new many-body dispersion corrected hybrid density functional, dubbed PBE0+MBD* by its original authors. For the purpose of validation, we first compare to published, high-level benchmark conformational energy hierarchies (coupled cluster at the singles, doubles, and perturbative triples level, CCSD(T)) for 73 conformers of small three-residue peptides, establishing that the van der Waals corrected PBE0 functional yields an average error of only ∼20 meV (∼0.5 kcal/mol). This compares to ∼40−50 meV for non-dispersion corrected PBE0 and 40−100 meV for different empirical force fields (estimated for the alanine tetrapeptide). For longer peptide chains that form a secondary structure, CCSD(T) level benchmark data are currently unaffordable. We thus turn to the *experimentally* well studied Ac-Phe-Ala$_5$-LysH$^+$ peptide, for which four closely competing conformers were established by infrared spectroscopy. For comparison, an exhaustive theoretical conformational space exploration yields at least 11 competing low energy minima. We show that (i) the many-body dispersion correction, (ii) the hybrid functional nature of PBE0+MBD*, and (iii) zero-point corrections are needed to reveal the four experimentally observed structures as the minima that would be populated at low temperature.



1. INTRODUCTION

The structure of a polypeptide or protein plays a central role for its function. At present much (essentially all) of the atomistic modeling and understanding achieved for systems of this size employs "force fields". These are rather simple analytical expressions of the potential energy in terms of the positions of the nuclei, in which the strengths of the considered interactions are parametrized. In practice, their reach is limited by two separate (but related) aspects: First, their necessarily limited functional form, and second, the fact that their parameters are obtained from a limited (even if large) base of experimental or theoretical input data. For instance, emphasis is often given to experimentally known conformation(s). The resulting potential-energy surface (PES) is thus (most likely) accurate at the energy minimum that was considered for determining the parameters, but for energy barriers and metastable states the reliability is in doubt. Indeed, many different force fields and parametrizations exist. Their Ramachandran plots, for example, reveal significant differences,[1,2] as do larger-scale calculated conformational properties (e.g., refs 3−6). Still, these parametrizations are currently the only methods fast enough to provide sufficient atomistic simulation data for an assessment of the (classical) statistical mechanics averages for systems comprising tens of thousands of atoms and more.

Obviously, the accuracy requirements to capture the PES of a peptide chain are rather stringent. Peptide chains are flexible and can assume many competing conformations. The structure (or structural ensemble) that they assume at given external conditions (temperature, pressure, environment) depends sensitively on the balance between different weak interactions. Very different conformations must be represented with the same systematic accuracy. In addition, processes like protonation and deprotonation of a chain, interactions with ions, bond formation, and bond breaking are aspects that can impact the conformational preference of a peptide, but that are difficult to capture in empirical formulations.

It would be ideal if one could switch over these simulations to quantum-mechanical first principles methods which have a wider range of validity. However, the highest-level theoretical

Special Issue: Kenneth D. Jordan Festschrift

Received: December 9, 2013
Revised: January 8, 2014
Published: January 9, 2014

   dx.doi.org/10.1021/jp412055r | J. Phys. Chem. A 2014, 118, 7349−7359

methods (e.g., coupled-cluster theory in the singles, doubles, and perturbative triples, CCSD(T)) are computationally so much more demanding that the system sizes which can be treated even for individual, fixed geometries are restricted to just a few amino acids at best.

Among the more affordable quantum-mechanical approaches, density-functional theory (DFT) is the most prominent candidate to provide the necessary balance between accuracy and efficiency. DFT is now frequently used for protein-related questions (see, e.g., refs 7−9 and references therein) but it is still far from trivial to establish which level of DFT provides truly reliable results. Even for simple peptide chains, small errors of, say 10 meV per residue, can completely alter the preferred structure of a given sequence of amino acids.[10−14]

The problem is exemplified particularly well by a recent series of experimental and theoretical studies of a seven-residue peptide in the gas phase, Ac-Phe-Ala_5-$LysH^+$, by the Rizzo group.[15−17] Using UV-IR double resonance spectroscopy and theoretical spectra calculated at the B3LYP/6-31G** level of theory, the coexistence of four distinct conformers (labeled A,B,C,D) at $T \approx 10$ K was convincingly established. The Rizzo group has recently shown these conformers to be different mixed 3_{10}-α helices,[16] and even their relative abundances were tentatively established. However, a follow-up computational study[17] involving the same group and employing 19 different semilocal and hybrid DFT functionals at the DZP level, as well as Hartree−Fock and second-order Møller−Plesset (MP2) theory, showed that none of these theoretical approaches produces a conformational energy hierarchy that matches the observed relative abundances of the four conformers.

Looking at the performance of DFT-based methods in general, noticeable deviations from reference data in energy hierarchies are well documented already for small hydrogen bonded systems (e.g., refs 18−24). Even simple water clusters treated with different functionals can exhibit deviations of more than 40 meV (1 kcal/mol) per water molecule, or per H-bond, from reference data.[20,21] Yet, conformational energy differences of this magnitude can be resolved in dedicated low-temperature experiments. Another documented case used as a benchmark in our own group[24] is the adsorption of two water molecules at a protonated valine amino acid. Here, an experimental infrared-spectroscopic study[22] (also by the Rizzo group) unambiguously identified a specific adsorption geometry. In contrast DFT-B3LYP favors a different geometry by ∼50 meV.[22,23] In a benchmark of our own,[24] the same (incorrect) qualitative ordering was found for two dispersion-corrected functionals,[25] dubbed PBE+vdW and PBE0+vdW, including zero-point vibrational effects. Clearly, careful assessments of and refinements to current DFT-based methods are in order to validate their reach for larger systems (peptides) with subtle conformational energy hierarchies of ∼100 meV or less.

In this paper, we focus on the performance of a group of dispersion-corrected density functionals in benchmarks for 73 conformers of tri- and tetrapeptides, as well as in the challenging case of the larger, secondary structure forming Ac-Phe-Ala_5-$LysH^+$. Specifically, we assess the widely used PBE[26] and PBE0[27] functionals including a state of the art pairwise-atomic dispersion correction[25] and a recent many-body van der Waals (vdW) dispersion correction[28,29] that we shall refer to as PBE(0)+vdW (pairwise) and PBE(0)+MBD* (many-body).[29] Both vdW corrections depend on the molecular geometry by way of the electronic density. In contrast to the pairwise approach, the MBD* variant can also capture the important nonadditive contribution originating from simultaneous dipole fluctuations at multiple atomic sites up to infinite order, leading both to an overall screening and long-range anisotropic effects. At the first-principles level, this effect is otherwise first accounted for in the much more expensive random-phase approximation (RPA).[30] These effects have been recently shown to be especially relevant for large molecular systems.[31] The MBD* approach employed here is a refined version of the MBD version published in ref 28. In the MBD* approach, the same group introduces an additional range-separation scheme into their expressions for locally screened effective polarizabilities and for the full many-body response (for details, see below, as well as a forthcoming paper by Ambrosetti and co-workers[29]).

We first compare to two sets of published, high-level benchmark data (CCSD(T)) for the conformational energy hierarchy of tri- and tetrapeptides: Gly-Phe-Ala (GFA), Gly-Gly-Phe (GGF), and Phe-Gly-Gly (FGG),[12] and Ac-Ala_3-NMe.[11,32] We then assess the isolated seven-residue peptide Ac-Phe-Ala_5-$LysH^+$ as a landmark experimental benchmark. As mentioned above, several DFT methods failed to explain the energy hierarchy of the A,B,C,D conformers of Ac-Phe-Ala_5-$LysH^+$ in a recent theoretical study.[17] One interesting observation is that all tested functionals[17] lack long-range van der Waals (vdW) dispersion contributions, which should have a large impact on conformational preferences.[33−35] However, the simultaneous occurrence of all four conformers in the same cryocooled ion trap does imply that they must be very close in energy. This would render their distinction difficult in any approximate computational method.

We also show that the real conformational problem of Ac-Phe-Ala_5-$LysH^+$ goes much deeper. In addition to the four experimentally established conformers, a 100-atom peptide chain exhibits a myriad of other possible structures. Thus, the criterion that a reliable theory must meet is not only to reproduce the conformational energies of conformers A, B, C, and D but also to predict all other possible but metastable conformers. Though they are less favorable, they may play a significant role in the dynamics.

In the following, we first give details about the computational methods used. We then focus on the conformational energy hierarchy of the tri- and tetrapeptides at the PES. Within the (very reasonable) error levels established for these smaller cases, we show that an exhaustive, first-principles conformational search of the PES, and low-temperature vibrational entropy effects for Ac-Phe-Ala_5-$LysH^+$ give a consistent picture at the most sophisticated level of theory investigated here, the many-body dispersion corrected hybrid functional PBE0+MBD*.

2. COMPUTATIONAL METHOD

All DFT calculations reported in this work were performed with the all-electron, localized-basis FHI-aims program package[36,37] and *tight* settings for the integrations grids and basis sets. These settings ensure that remaining numerical and basis-set errors amount to only a few meV per H-bond for energy differences.[24,36]

Both dispersion corrections analyzed in this work are used as implemented in FHI-aims by their original authors. The pairwise van der Waals dispersion correction ("+vdW") was described in ref 25. The effective atomic polarizabilities used to determine the pairwise correction depend on the geometry and

are derived from the self-consistent electron density of the full molecule or solid.

As mentioned, the many-body dispersion correction ("+MBD*") investigated below is a refined version of the type originally reported in ref 28. It is *not* a simple sum of pairwise or higher-order terms. Instead, in this approach, dispersion interactions are modeled as coupled quantum harmonic oscillators located at individual atomic sites. The oscillator polarizabilities and frequencies are obtained by first applying a Dyson-like self-consistent screening correction to the bare effective polarizabilities determined according to ref 25. In a second step, the screened polarizabilities and frequencies are inserted into a system-wide coupled fluctuating dipole model Hamiltonian that is then solved by diagonalization to give the dispersion energy. Intriguingly, this approach can be shown[30] to be equivalent to the RPA expression for the model dipole system. The frequently used pairwise dispersion energy approach is formally equivalent to the second-order term of the RPA expression.

The refinement,[29] called "MBD*", is achieved by range-separating the dipolar interaction tensor defined in ref 28. Only its short-range part is used to define the needed screened oscillator polarizabilities and frequencies. Conversely, only its long-range part is then used to evaluate the full coupled fluctuating dipole model Hamiltonian. This guarantees an expression free of double-counting effects. A full description will be given by the original authors of the "MBD*" method in a forthcoming paper.[29]

Either correction is always coupled with a standard semilocal or hybrid density functional, leading to the following notation:

- PBE+vdW for the PBE functional[26] with the pairwise vdW correction[25]
- PBE+MBD* for the PBE functional with the refined many-body correction,[29] MBD*
- PBE0+vdW for the PBE0 hybrid functional[27] with the pairwise vdW correction[25]
- PBE0+MBD* for the PBE0 functional and MBD* correction[29]

For all zero-point and harmonic finite-temperature free-energy corrections, we use the vibrational frequencies calculated at the level of DFT-PBE+vdW in the harmonic approximation. These frequencies are numerically converged to an accuracy better than 2 cm^{-1} (see ref 14 for an explicit benchmark). We do *not* recompute the frequencies at either the PBE0 level or the MBD* level. For the PBE0 functional, the computational cost of a fully converged finite difference approach for several conformers in a 100-atom system at high accuracy (*tier2* basis set and *tight* settings) is presently still excessive, even in an overall linear-scaling implementation of the Hartree−Fock exchange term in FHI-aims. For the MBD* functional, analytic gradients (required in the finite difference implementation) are not yet included in the early implementation available to us.

We also estimate the rotational free energies within the rigid rotor approximation, which yield energy differences of no more than 10 meV. We add this correction also to the PBE0(+vdW, +MBD*) energies.

3. RESULTS AND DISCUSSION

3.1. Benchmarks: Tri- and Tetrapeptides. We first analyze the performance of the PBE semilocal exchange-correlation functional and the PBE0 hybrid exchange correlation functional for prototype peptides that are relevant for this work. We test the standard functionals, as well as the functionals with the pairwise and many-body dispersion corrections.

The peptides tested are different conformations of the GFA, GGF, FGG, and Ac-Ala_3-NMe (alanine tetrapeptide) peptides. Conformational energy hierarchies computed at the CCSD(T) level were reported for the first three peptides in ref 12 and for the last in the Supporting Information of ref 32. In both references the CCSD(T) energies were obtained by summing the complete basis set limit extrapolated MP2 energy and the difference between the CCSD(T) and MP2 energies at the 6-31G* basis sets. As these small-peptide studies are performed for comparing and validating different total-energy methods, we take the same geometries used in refs 12 and 32. To illustrate the broad variation of possible conformers even for these rather short peptide chains, we show all geometries for the four different peptides as used in the present paper in Figure 1. It is this variety that a reliable theoretical description of their PESs should describe accurately.

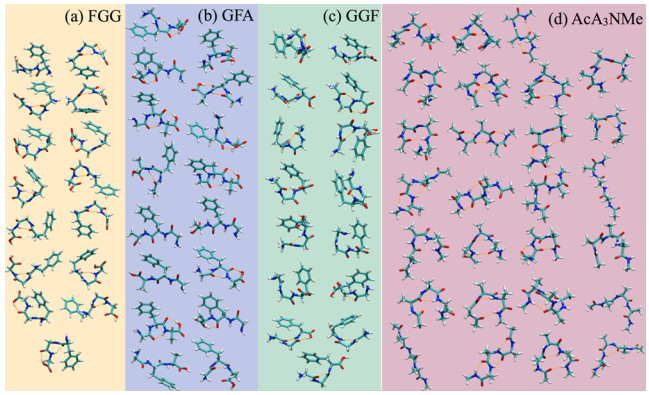


Figure 1. Different conformations of the (a) FGG, (b) GFA, (c) GGF, and (d) Ac-Ala_3-NMe peptides used in the benchmark assessment presented in Table 1

In Table 1, we report the observed deviations of the DFT-based calculated energy hierarchies with respect to the CCSD(T) reference data of refs 12 and 32. We computed the mean absolute error (MAE), defined as

$$MAE = \frac{1}{N}\sum_{i}^{N} |\Delta E_i - \Delta E_i^{CCSD(T)} + b| \quad (1)$$

where i runs over all N conformations in a given set. ΔE_i corresponds to an energy difference between conformation i and a chosen reference conformation. To ensure that the eventual MAE is independent of the reference conformation, b is a constant parameter ("offset") adjusted so as to minimize the MAE. In practice, b was determining by finding the best fit of the data to a curve of the form $y = x + b$. We also show in Table 1 the maximum absolute error MAX, defined as $\max_i[|\Delta E_i - \Delta E_i^{CCSD(T)} + b|]$.

Three conclusions can be drawn from Table 1. (i) For both the PBE and PBE0 functionals, the dispersion corrections substantially improve the performance of the functionals. (ii) The dispersion corrected (both pairwise and many body) PBE0 functional shows a superior performance compared to the dispersion corrected PBE functional in systems containing phenylalanine. In contrast, both functionals show a roughly

dx.doi.org/10.1021/jp412055r | J. Phys. Chem. A 2014, 118, 7349−7359

Table 1. Mean Absolute Error (MAE) and Maximum Error (max) for the Energy Hierarchies of 16 Conformers of Gly-Phe-Ala (GFA), 15 Conformers of Gly-Gly-Phe (GGF), 15 Conformers of Phe-Gly-Gly (FGG), and 27 Conformers of Ac-Ala$_3$-NMe, Compared to CCSD(T) Reference Data from refs 11 and 12. Energies are Reported in meV (in Parentheses: Converted to kcal/mol)

	PBE	PBE+vdW	PBE+MBD*	PBE0	PBE0+vdW	PBE0+MBD*
GFA						
MAE	53(1.2)	32(0.7)	44(1.0)	40(0.9)	17(0.4)	25(0.6)
max	108(2.5)	88(2.0)	76(1.7)	89(2.0)	72(1.7)	61(1.4)
GGF						
MAE	48 (1.1)	36(0.8)	40(0.9)	38(0.9)	26(0.6)	28(0.6)
max	143(3.3)	99(2.3)	84(1.9)	119(2.7)	78(1.8)	66(1.5)
FGG						
MAE	43(1.0)	37(0.8)	36(0.8)	35(0.8)	23(0.5)	23(0.5)
max	160(3.7)	59(1.4)	88(2.0)	132(3.0)	38(0.9)	59(1.4)
Ac-Ala$_3$-NMe						
MAE	55(1.3)	21(0.5)	22(0.5)	54(1.2)	18(0.4)	20(0.5)
max	131(3.0)	72(1.7)	66(1.5)	132(3.0)	47(1.1)	54(1.2)

	OPLS-aa	Amber99sb	Charm22	AmoebaPro04
Ac-Ala$_3$-NMe				
MAE	108(2.5)	42(1.0)	91(2.1)	53(1.2)
max	246(5.7)	86(2.0)	271(6.2)	112(2.6)

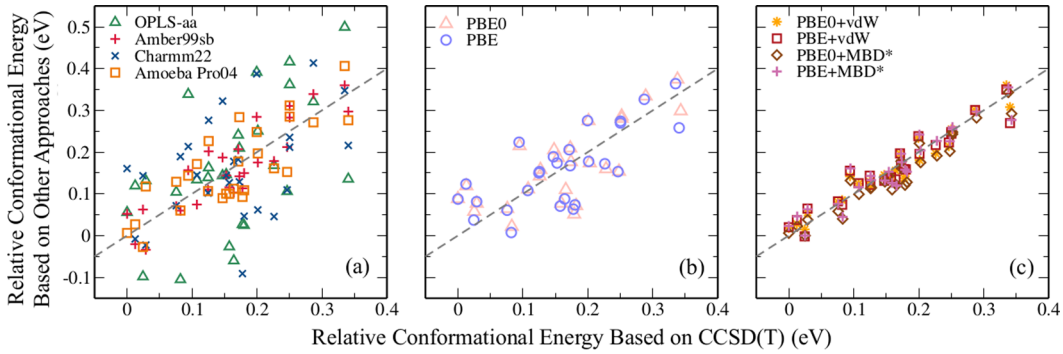


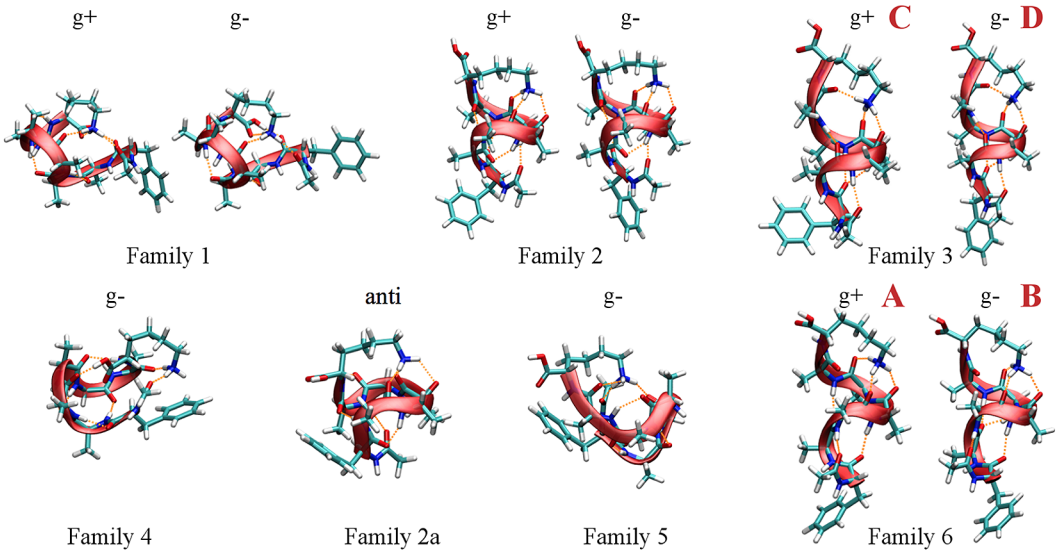
Figure 2. Comparison of the CCSD(T) conformational energy hierarchy of 27 Ac-Ala$_3$-NMe[12,32] to conformational energy hierarchies obtained by other computational methods for the same geometries in this work. In each graph, symbols indicate conformers ordered along the *x*-axis according to their CCSD(T) energy, relative to a single reference conformer. The *y*-position of symbols for each conformer indicates the conformational energy hierarchies according to different benchmarked computational methods, distinguished by different symbols in the plot. The dependence on any single reference structure is eliminated according to eq 1. The dashed diagonals indicate hypothetical perfect agreement between CCSD(T) and other methods. (a) Standard OPLS-aa, Amber99sb, Amoeba-Pro04, and Charmm force-fields, (b) standard PBE and PBE0 functionals, and (c) dispersion corrected PBE+vdW, PBE+MBD*, PBE0+vdW, PBE0+MBD* (see text).

similar performance for the alanine tetrapeptide. This effect is possibly related to the better performance of hybrid functionals especially for multiply-bonded systems[38] and to the better performance of PBE0 at predicting static polarizabilities of several systems, including aromatic structures.[39] (iii) The comparison of the pairwise and the many-body dispersion correction does not reveal a large difference, with deviations within the uncertainties of the reference data and the calculations. For the relatively small systems investigated here, the approximate equivalence of the pairwise and the many-body dispersion scheme is in fact not surprising, since the full impact of the coupled harmonic oscillator description in MBD* will become larger as the system size increases. We note the small change of 8 meV of the MAE for the alanine tetrapeptide using PBE0+MBD* compared to the MAE reported for the original "MBD" version of ref 28.

For the alanine tetrapeptide we also analyzed the performance of several standard empirical force fields, which we show in Figure 2 and the second part of Table 1. Using the TINKER package,[40] we calculated the relative energies of the conformers with the OPLS-aa, Amber99sb, Charmm (nonpolarizable), and AmoebaPro04 (polarizable) force fields. What we find is that the two best-performing tested force fields, Amber99sb and AmoebaPro04, show a performance which is similar to the standard PBE and PBE0 functionals (no dispersion corrections) for this particular system. In this context, it is interesting to note that a set of 51 alanine tetrapeptide conformations entered the construction of Amber99sb,[41] which may partially explain the good performance of Amber99sb. OPLS-aa and Charmm perform significantly worse. While especially the Amber99sb and AmoebaPro04 performance is a significant success, the dispersion-corrected density functionals PBE0+vdW and PBE0+MBD* still yield less than half the MAE for Ac-Ala$_3$-NMe.

To translate the errors reported here to the conformational energy hierarchy of the one hundred atom system Ac-Phe-Ala$_5$-LysH$^+$, one must account for different system sizes. The maximum error in our benchmarks is of the order of 50 meV for 40 atoms (PBE0+vdW or PBE0+MBD*, alanine tetrapeptide). This implies that it could conceivably reach 125 meV for

dx.doi.org/10.1021/jp412055r | J. Phys. Chem. A 2014, 118, 7349−7359



ΔE (eV)	Family 1		Family 2		Family 3		Family 4	Family 5	Family 2a	Family 6	
PES	*g*+	*g*-	*g*+	*g*-	*g*+ **C**	*g*- **D**	*g*-	*g*-	*anti*	*g*+ **A**	*g*- **B**
PBE	0.00	0.02	0.03	0.00	-0.04	-0.08	0.30	0.22	0.13	0.01	-0.03
PBE+vdW	0.00	0.10	0.04	0.07	0.07	0.10	0.10	0.12	0.13	0.13	0.15
PBE+MBD*	0.00	0.09	0.05	0.07	0.08	0.08	0.14	0.18	0.16	0.11	0.11
PBE0	0.00	0.03	0.00	-0.04	-0.09	-0.12	0.29	0.24	0.12	-0.06	-0.09
PBE0+vdW	0.00	0.10	-0.01	0.02	0.01	0.03	0.10	0.12	0.12	0.05	0.06
PBE0+MBD*	0.00	0.08	0.01	0.03	0.03	0.02	0.14	0.18	0.14	0.03	0.03

Figure 3. Representative conformers and conformational energy hierarchies of the lowest energy H-bond families of Ac-Phe-Ala$_5$-LysH$^+$. Families 1, 2a, 2, 4, and 5 are nonhelical while Family 3, 6 are helical. Conformational energies for the DFT-PBE, -PBE0, -PBE+vdW and -PBE0+vdW functionals are given for fully relaxed local minimum structures of the respective potential-energy surfaces (PES). For the −MBD* approach, the local structure optima of the pairwise ("−vdW") dispersion corrections were used. The *g*+ and *g*− conformers of Families 6 and 3 are labeled A, B, C, and D, respectively. This follows the notation of ref 15, which reports A, B, C, D as the best matches to conformer-specific, low-temperature experimental IR spectra. Detailed H-bond patterns are given in Table 2

Ac-Phe-Ala$_5$-LysH$^+$, albeit the situation will likely be better for the structurally similar conformers A, B, C, and D. For the MAE, we expect much smaller values (around 50 meV).

The above error estimates are central when assessing the cost−benefit ratio of DFT based methods for the peptide chain. With present methods, one can certainly expect a significant improvement over even the best empirical approaches. What cannot (yet) be expected for a peptide chain of 100 atoms or more is accuracy below (say) 1 kcal/mol for individual conformational energy differences. However, perhaps the most significant benefit of density-functional based methods is the expected absence of uncontrolled systematic errors across very different regions of conformation space. This property is very important for the correct evaluation of statistical mechanics ensemble properties.

3.2. Ac-Phe-Ala$_5$-LysH$^+$: Conformational Search. The conceivable conformational space of the 100-atom peptide Ac-Phe-Ala$_5$LysH$^+$ is formidably large. To search it thoroughly for low-energy conformers, we use the strategy already employed in refs 14 and 24. We begin with an initial enumeration of possible candidate geometries by performing a basin-hopping search as implemented in the TINKER package,[40] using the OPLS-aa[42] force field (FF). The search follows the 15 softest torsional mode directions from each newly identified PES minimum. During the search, any new local minima 50 kcal/mol or above the current lowest-energy minimum are discarded. The complete procedure identifies 282,022 PES minima at the FF level. Given the sheer number of structures explored, we do not expect significantly different overall results by changing the force field for the initial enumeration.

We next fully relax the 1000 lowest-energy FF geometries at the PBE+vdW level of theory, starting with FHI-aims *light* settings for integration grids and basis sets. From these 1000 conformers, we take the 60 lowest-energy (PBE+vdW) conformers and fully relax them with *tier2* basis sets and *tight* numerical settings.[36] In this final step (from *light* to *tight*), the energy hierarchy changes are only of the order of 10 meV.[14] The 60 conformers span an energy window of 0.18 eV, based on the PBE+vdW *tight* energy hierarchy.

The 60 structural minima were classified into families according to their hydrogen bond patterns. A hydrogen bond was considered to be present if the distance between the donor CO group and the acceptor NH group was less than 2.5 Å. Here, members of a particular family may differ in the exact orientation of the LysH$^+$ side chain, or the COOH group near the C-terminus, or the orientation of the phenylalanine side chain. The Phe χ angle, measured along the N−C$^\alpha$−C$^\beta$−C$^\gamma$ atoms, may have values of approximately 180, +60, or −60° corresponding to labels *anti*, *gauche*+ (*g*+) and *gauche*− (*g*−), like in ref 15.

Finally, we take the lowest energy PBE+vdW *tight* conformer of each family, with all identified different orientations of the phenyl ring (a total of 11 conformers, shown in Figure 3), and fully relax them with the PBE0+vdW functional, with *tight* numerical and basis set settings. For the "MBD*" many-body dispersion corrected functionals, we use the local structure minima of the pairwise approach ("+vdW") and only perform single-point energy evaluations. This choice is justified since the overall total energy changes between the pairwise and the MBD* approaches (which could affect the local PES minimum

dx.doi.org/10.1021/jp412055r | J. Phys. Chem. A 2014, 118, 7349−7359

Table 2. H-Bond Network of the Seven Lowest Energy Families of Ac-Phe-Ala$_5$-LysH$^+$[a]

	Family 1	Family 2	Family 2a	Family 3	Family 4	Family 5	Family 6
O(Ac)	NH_3^+	NH5(π)	NH5 (π) NH6	NH3(3_{10})	NH_3^+	NH_3^+	NH3(3_{10})
O1	COOH	NH6(π)	NH3(2_7)	NH4(3_{10}) NH5(α)	NH3(2_7)	NH3(2_7)	NH4(3_{10})
O2	NH4(2_7)	NH4(2_7)	NH_3^+	NH6(α)	NH4(2_7) NH5(3_{10})	NH5(3_{10})	NH5(3_{10}) NH6(α)
O3	NH6(α)	NH_3^+	NH_3^+	NH_3^+	COOH	NH_3^+	NH_3^+
O4	NH_3^+	NH_3^+	NH_3^+	NH_3^+	NH(Lys)	NH6(2_7) NH_3^+	NH_3^+
O5	NH(Lys)	NH_3^+	NH7 (2_7)	free	NH2 (inv.)	NH2 (inv.)	NH_3^+
O6	NH3 (inv.)	NH3 (inv.)	NH2 (inv.)	NH_3^+	NH_3^+	NH_3^+	free
COOH	NH_3^+	free	free	free	free	free	free

[a]The first column indicates to which residue (or capping) the oxygen in question belongs. For each family, the other columns indicate the position of the hydrogen to which this oxygen is hydrogen-bonded.

structures) are small compared to typical energy changes encountered for appreciable geometry changes of a 100-atom molecule.

3.3. Ac-Phe-Ala$_5$-LysH$^+$: Potential Energy Surface. We first characterize the structure and conformational energy hierarchy of the 11 final conformers (Figure 3) at the PES only, i.e., without taking into account rotational and vibrational free-energy contributions. As mentioned, we find six hydrogen-bond "families" within 0.15 eV of the lowest-energy structure. The corresponding H-bond patterns are shown in Table 2. These hydrogen bonds are also indicated as dashed orange lines in Figure 3. We note that the hydrogen bonds of the α-helix define a characteristic ring of 14 atoms in the structure. For the 3_{10}-helix, the equivalent ring has 10 atoms. On the basis of the presence or absence of such rings, we classify five of our families as nonhelical (Family 1, Family 2, Family 2a, Family 4, and Family 5). The other two (Family 3 and Family 6, each with $g+$ and $g-$ oriented phenyl rings) are the mixed helices already identified experimentally[15] as conformers A, B, C, and D. For brevity, we denote the families and their orientation of the phenyl ring by $fN\pm$, where N goes from 1 to 6.

Family 1 is a compact turn structure that could be characterized as an α-turn. Family 2a is substantially compact, binding its central carbonyl groups directly to the NH_3^+ group. Family 2 is a structure similar to the conformer named "g-1" in ref 33, having characteristic π-helical loops and an "inverted" H-bond pointing against the backbone loop dipole. Family 3 is a mixed $3_{10}/\alpha$ helix that corresponds to the conformers C and D of ref 15, with the $f3+$ and $f3-$ conformations, respectively. Families 4 and 5 are compact β-turn structures, characterized by the presence of a 3_{10}-like loop. Family 6 is another mixed $3_{10}/\alpha$-helix that, in the $f6+$ and $f6-$ conformations, corresponds to conformers A and B of ref 15.

In Figure 3, we also report the energy hierarchies on the PES of the PBE, PBE+vdW, PBE+MBD*, PBE0, PBE0+vdW, and PBE0+MBD* functionals (no vibrational or rotational free-energy contributions yet). We first use them to highlight differences between the different functionals; for a comparison to experiment, free-energy contributions are discussed further below.

Consistent with ref 17, we do see discrepancies between the PBE and PBE0 functionals at all levels. The numbers reported in ref 17 for the PBE and PBE0 functionals (conformers A, B, C, and D) are slightly different from ours because of different computational protocols, especially the rather small DZP basis sets used in ref 17.

Beginning with PBE and PBE0 (not dispersion-corrected), it is evident that the change from the semilocal (PBE) to hybrid (PBE0) functional changes the hierarchy slightly but systematically. In particular, the two helical families 3 and 6 are stabilized by the hybrid functional. Changes of this magnitude are consistent with the benchmarks for smaller peptides shown in Table 1.

Looking back at the smaller-peptide benchmarks of Table 1, it is clear that dispersion interactions are essential to reflect the PES adequately (as is well-known in the literature[43]). In Figure 3, we see that the inclusion of long-range dispersion interactions for Ac-Phe-Ala$_5$-LysH$^+$ has an even larger effect than for the smaller peptides, in some cases changing the energy hierarchy by close to 100 meV (see Family 6, for example). It is also evident that more compact conformers (the nonhelical ones) are more stabilized compared to the more extended, helical ones. Likewise, the changes are also consistently more pronounced for the $g-$ members of each family. In parts, this also affects the geometry of the phenyl ring. While the root-mean-square deviation between the relaxed structures with PBE(0) and PBE(0)+vdW does not exceed 0.6 Å (see Supporting Information), the phenyl ring always tends to get closer to the backbone in the "+vdW" functionals.

Regarding the differences between the two dispersion treatments, pairwise ("+vdW") or many-body (MBD*), the conformational energy hierarchies change noticeably by a few tens of meV. The changes are clearly larger than for the smaller peptides (Table 1). In particular, the helical families $f3+$ and $f3-$ (C and D of ref 15, respectively) and $f6+$ and $f6-$ (A and B of ref 15, respectively) are much closer in energy in the MBD* treatment than in the pairwise treatment. Interestingly, the PBE0+vdW (pairwise) conformational energy hierarchy C < D ≈ A < B is close to the MP2 conformational energy hierarchy of ref 17. Even more intriguingly, in PBE0+MBD*, all four become essentially isoenergetic. However, at the PES, the helical conformers are still only found within the lowest 50 meV of the conformational energy hierarchy, even in the PBE0+MBD* case. They compete with the nonhelical $f1+$, $f2+$, and $f2-$, respectively.

3.4. Ac-Phe-Ala$_5$-LysH$^+$: Impact of Vibrational and Rotational Free-Energy Contributions. Experimentally,[15,16] only the conformers A, B, C, and D are observed, meaning that they were the only conformers populated at the experimental

dx.doi.org/10.1021/jp412055r | J. Phys. Chem. A 2014, 118, 7349−7359

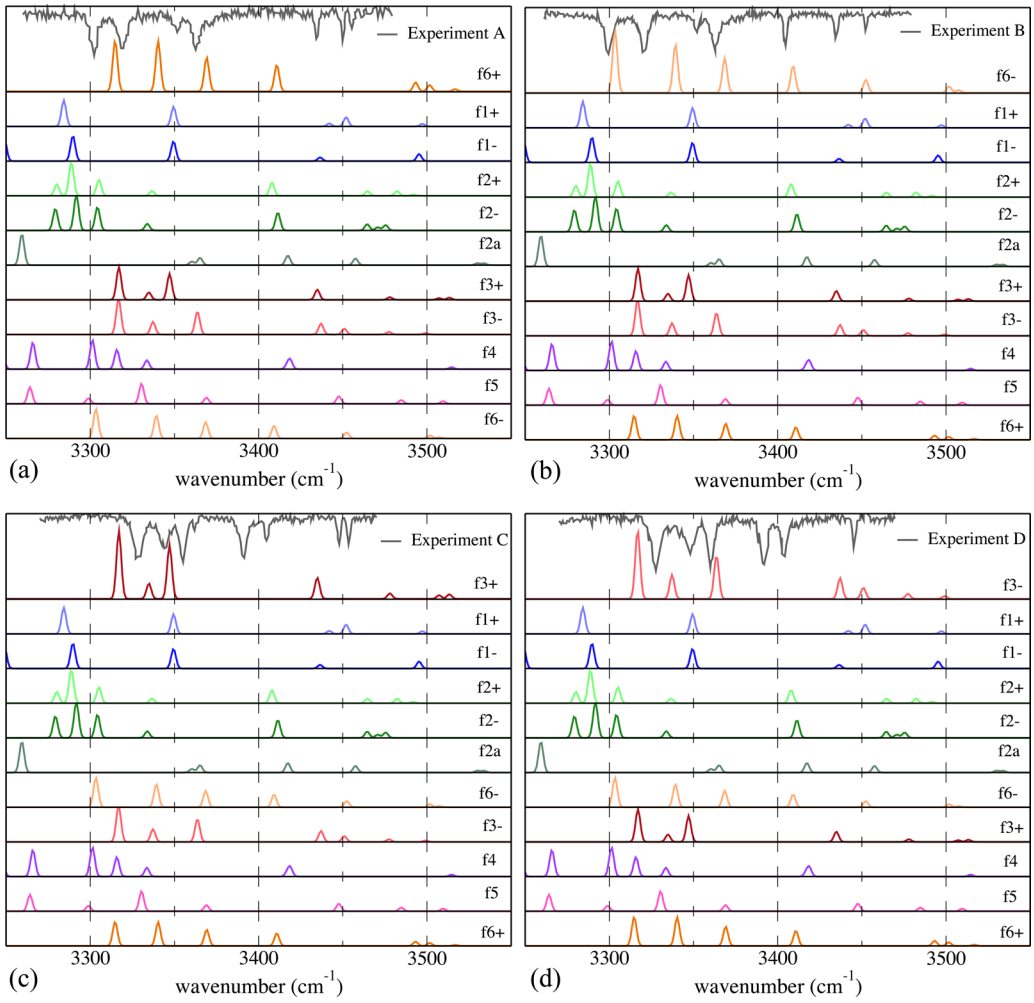


Figure 4. Harmonic IR-spectra of Ac-Phe-Ala$_5$-LysH$^+$: The top panel in each plot shows the experimental spectra for one of the four conformers (a) A, (b) B, (c) C, and (d) D, as proposed in refs 15 and 16. The second panel below this is the theoretical harmonic vibration spectra of that specific conformer, calculated using the PBE+vdW functional and "tight" settings in FHI-aims. The lower panels show the spectra of the other low energy conformers found during our conformer search. All calculated spectra have been convoluted with a Gaussian of 4 cm^{-1} fwhm for better visual comparison.

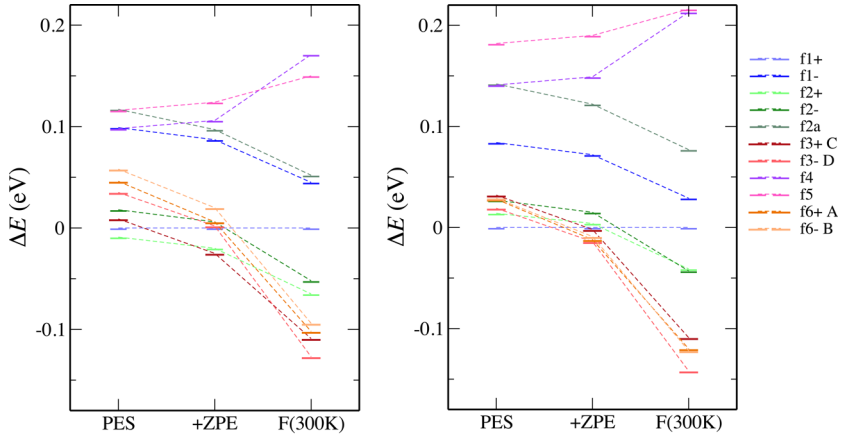


Figure 5. Energy hierarchies of the Ac-Phe-Ala$_5$-LysH$^+$ conformers presented in Figure 3. In panel a, the hierarchy for the PBE0+vdW total energies is presented in the first column, the ZPE corrected one is shown in the second column, and the one adding harmonic free energy contributions at 300 K is shown in the third column. In panel b we show the same as in panel a for the PBE0+MBD* functional, which includes many-body van der Waals contributions.

dx.doi.org/10.1021/jp412055r | J. Phys. Chem. A 2014, 118, 7349−7359

conditions (at low temperature, $T \approx 7$ K). The primary conformation-sensitive evidence in refs 15 and 16 are conformer-selected infrared spectra (the conformation selectivity is given by IR-UV double-resonance spectroscopy). In the original work, comparisons to DFT-B3LYP computed spectra demonstrated the qualitative correspondence of conformers A, B, C, D to the respective IR spectra.

To corroborrate the assignment of refs 15 and 16 we compare the harmonic vibrational frequencies and intensities of all our 11 conformers at the DFT-PBE+vdW level to experiment. The result, for all 11 conformers of Figure 3, is shown in Figure 4. It turns out that this comparison can only remain at a qualitative level due to the well-known difficulties of semilocal DFT and the harmonic approximation in the hydrogen stretch region. Empirical procedures to partially rectify this are often employed (e.g., refs 44 and 45 and many others), but we chose not to apply any empirical scaling procedure to our spectra. What Figure 3 confirms is that there is no other conformer than A, B, C, and D that would provide a better match to the experimental data. That said, the overall comparison between the experimental and calculated (harmonic, PBE+vdW) spectra in this frequency range is disappointing. We expect that *both* improvements to the density functional and a proper (quantum) anharmonic treatment are needed to yield unambiguously improved theoretical spectra. Although very expensive, we have also recomputed the spectrum for conformer C ($f3+$) with DFT-PBE0+vdW and with a reduced basis set. The details are shown in the Supporting Information. Even if all peaks lie much higher in frequency for this functional (a known feature of hybrid functionals[44,45]) the match to experiment is not bad (but also not better than DFT-PBE +vdW) after applying a uniform scaling factor, as is often done in the literature.

In Figure 5, we show the predicted conformational energy hierarchy for Ac-Phe-Ala_5-$LysH^+$ without and with vibrational free-energy effects. For the PES, we focus on the PBE0+vdW and PBE0+MBD* functionals, which clearly emerged as the superior choices in Table 1. Data equivalent to Figure 5 for the PBE+vdW and PBE+MBD* methods can be found in the SI.

As mentioned in section 2, we compute the vibrational free-energy part at the level of DFT-PBE+vdW and *tight* settings (*tier2* basis sets), due to the much larger computational cost of the DFT-PBE0+vdW at the *tier2* basis set level.

By adding the zero-point contribution, only $f4-$ and $f5-$ get destabilized with respect to $f1+$. The other conformers are stabilized with respect to $f1+$ and change their energy ordering. Indeed, the PBE0+MBD* method predicts the A, B, C, and D conformers to be the lowest free-energy structures already at $T = 0$ K in the harmonic approximation, with all four experimental conformers within a few meV of one another. Focusing on those four conformers first, we do expect the four observed conformers to lie extremely close in energy, since they should be all accessible in the conformational ensemble at a given temperature, and even at room temperature k_BT is only 26 meV. Although technically within the error limits that we can assign to the DFT-based methods employed here, we do note that the many-body dispersion contribution plays a significant role in bringing all four conformers energetically close together.

We also note that experimentally, A and B are suggested to be more abundant than C and D. This means that the first two conformers should be lower in energy, a key point pursued in ref.[17] In our results, we do not find this precise energetic ordering of A,B,C,D, similar to what was observed in ref 17. However, conformers A and B are practically isoenergetic with conformer D. With all its remaining uncertainties, the PBE0+MBD* functional indeed comes closest by far among all previously investigated DFT approaches to explaining the experimentally observed abundances. It is thus tempting to conclude that this exact ordering is indeed influenced by many-body effects that are otherwise captured only in much higher-level theories (RPA, CCSD, or beyond). We also note that the smaller-peptide benchmarks do indicate that we cannot expect our method to resolve these small conformational energy differences completely. Simply finding the four relevant conformers close in energy and as the lowest-free energy conformers overall already reflects a considerable success of the DFT-PBE0+MBD* approach.

Looking at the overall ZPE-corrected conformational energy hierarchy, $f1+$, $f2+$, and $f2-$ are still very close in energy to A,B,C,D. If this result were exactly correct, one might expect to also find those nonhelical conformers in experiment, but they are not reported.

A possible explanation is provided by the finite-temperature free-energy difference of the conformers in question. As shown in Figure 5, raising the temperature to around room temperature (in the harmonic approximation) unambiguously renders the four observed structures the isolated lowest energy ones. The stabilization is consistent with the previous observation that helical structures are stabilized by vibrational entropic contributions with respect to more compact ones.[14] Although the experiments of refs 15 and 16 were nominally conducted at low temperature, the observed conformational ensemble was obtained after cooling from high temperature. It is thus possible that the ensemble would reflect a conformational mix frozen at somewhat higher temperature.[46] The experimental prevalence of the helical conformers would thus be a result of their vibrational entropic stabilization.

3.5. Computational Cost. For the cases studied here, the PBE0+MBD* functional qualitatively improves over semilocal density functional theory and over earlier, pairwise dispersion correction schemes. Importantly, these improvements are physically motivated, that is, this is not just a case of "shopping around for a different functional".

What is also important, however, is the fact that the Hartree−Fock like exchange operator in the hybrid functional drives up the computational cost. For the well converged "tight" settings and the Ac-Phe-Ala_5-$LysH^+$ molecule studied here, the time increase from PBE to PBE0 is a factor of 40−50 in the presently used implementation, that is, a significant expense. In contrast, neither the pairwise nor the MBD* approach to dispersion interactions contribute significantly to the computational cost.

We have thus also investigated a similar, but expected less costly functional, the Heyd-Scuseria-Ernzerhof[47] (HSE) hybrid functional in its 2006 version (HSE06),[48] combined with the MBD* dispersion term. The HSE06 functional employs a short-range screened exchange operator, the screening parameter of which was set to $\omega = 0.2$ Å^{-1}, as recommended in ref 48. Conformational energy hierarchy results analogous to Figure 5 for Ac-Phe-Ala_5-$LysH^+$ are also included in the Supporting Information, showing a small change of the critical conformational energy hierarchy between helical and nonhelical conformers by ∼20 meV.

Interestingly, for the system size investigated here, the use of the screened exchange operator in HSE06+MBD* did not

result in a significant reduction in computing time. The present (screened) exact exchange implementation in FHI-aims is already a sophisticated linear-scaling implementation,[49] similar in spirit to ref 50. This implementation makes use of the decay both of the Coulomb operator and the system's one-particle density matrix with distance; that is, a range reduction of the Coulomb operator should have an effect. The fact that it does not implies that (i) even a hundred-atom system may still be too small to feel the effect of the range-reduction at $\omega = 0.2$ Å^{-1}, (ii) already the decay of the density matrix may cover some of the intended effects of the range-reduced Coulomb operator.

The point is that to fully leverage the accuracy improvements promised by PBE0+MBD* or similar hybrid functionals for peptides, more efficient computational algorithms are clearly of interest. The available opportunities (ranging from more refined internal thresholds all the way to new hardware platforms such as GPU architectures) are certainly not yet exhausted, and are an ongoing focus of our groups.

4. CONCLUSION

We have assessed the performance of four dispersion-corrected density functionals, DFT-PBE+vdW, PBE+MBD*, PBE0+vdW, and PBE0+MBD*, with respect to their ability to correctly predict the conformational energy hierarchies of peptide chains. The investigated peptides reach from tri- and tetrapeptides to a more complex chain Ac-Phe-Ala$_5$-LysH$^+$, an experimentally well studied peptide that forms helical secondary structure.

With respect to the capability of the investigated methods, we conclude as follows:

1. By using a dispersion-corrected version of the PBE0 hybrid functional, the conformational energy hierarchy of tri- and tetra-peptides is reproduced with a mean absolute error of 20−30 meV.
2. Dispersion corrections are essential to achieve this accuracy. For short peptides, the difference between the pairwise ("+vdW") and the many-body dispersion ("+MBD*") treatment is rather small.
3. A set of standard protein force fields yields significantly higher MAEs (40−100 meV) for the tetrapeptide conformational energy hierarchy.
4. The experimentally observed prevalence of helical conformers for the longer peptide Ac-Phe-Ala$_5$-LysH$^+$ is reproduced by the dispersion-corrected DFT-PBE0+MBD* functional. In the case of this longer peptide, the many-body dispersion correction makes a noticeable difference.
5. Including the vibrational entropy is essential to recover the experimental energy hierarchy. Importantly, the vibrational contribution counters the effect of the dispersion correction, which on its own would favor more compact conformers.
6. Compared to the study of many more functionals in ref 17, the PBE0+MBD* functional including ZPE corrections performs best by far to explain the approximate abundances of conformers A, B, C, D inferred from experiment. The exact energetic ordering of the A, B, C, D conformers is still difficult to quantify, since they appear within a few meV of one another. Still, this small energy spread is remarkably consistent with their simultaneous presence in experiment at low temperature. What do these results imply for the prospects of density-functional based methods for peptide structure?

Most importantly, our results show the ability of a state of the art density functional to reproduce the conformational energy hierarchy of a one hundred atom peptide with an accuracy of ∼1−2 kcal/mol. This is a significant achievement, albeit accomplished by a still rather expensive hybrid functional (PBE0+MBD*). Although "purely quantum" simulations of large peptides and their statistical mechanics at these levels are thus somewhat in the future, there is no intrinsic qualitative obstacle: In PBE0, the effort to compute the most expensive part of the functional, the Hartree−Fock-like exact exchange contribution, scales as $O(N)$.

Our findings also reflect the nagging insight that simply putting a standard, so-called "first-principles" approach to work does not automatically amount to a reliable description, compared to the best empirical approaches. For the alanine tetrapeptide, the Amber99sb and AmoebaPro04 force field do rather well. As we show, it is still possible to do significantly better with current density-functional approximations for the right (physical) reasons: the inclusion of van der Waals dispersion interactions including their many-body nature, the inclusion of vibrational free energy contributions, and the mitigation of electronic self-interaction errors in the hybrid functional.

Leaving away the dispersion term from DFT or ignoring zero-point and finite-temperature vibrational effects would indeed significantly alter the conformational energy hierarchy, as would (in the present case) the description by the semilocal PBE functional part instead of PBE0.

In general, a big challenge to the field remains the adequate reflection of nuclear quantum effects in hydrogen-bonded systems for statistical simulations beyond the harmonic approximation. Perhaps the strongest effort to rectify this situation is made in the context of path-integral molecular dynamics (PIMD), although the cost and reach of these approaches is still far behind what can be accomplished in classical molecular dynamics today.

Nevertheless, the functionals investigated here imply significant progress toward direct, first-principles-based predictions of the properties of large and complex peptide chains. The availability of clean, unambiguous experimental benchmarks such as those published for Ac-Phe-Ala$_5$-LysH$^+$ is essential to develop current computational theory to a point where a fully reliable, first-principles picture of these important chemical and biological systems can be achieved.

■ ASSOCIATED CONTENT

Supporting Information

The SI contains the geometries for the structures of all conformers of Ac-Phe-Ala$_5$-LysH$^+$ discussed here, RMSD estimates between structures relaxed with and without vdW contributions, the PBE0+vdW spectrum of $f3+$, and the equivalent of Figure 5 for the PBE+vdW and PBE+MBD* functionals, as well as the HSE06+vdW and HSE06+MBD* functionals. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: mariana.rossi.work@gmail.com; volker.blum@duke.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank Prof. Thomas Rizzo and Caroline Seaiby for providing their original experimental data reproduced in Figure 4, and geometries for their calculated structures, that we could compare to ours. The authors also thank Alexandre Tkatchenko, Alberto Ambrosetti, and Vivekanand Gobre for making the code and manuscript for the MBD* dispersion correction available to us at an early date.

■ REFERENCES

(1) Penev, E.; Ireta, J.; Shea, J. Energetics of infinite homopolypetide chains: A new look at commonly used force fields. *J. Phys. Chem. B* **2008**, *112*, 6872.

(2) Best, R. B.; Buchete, N.-V.; Hummer, G. Are Current Molecular Dynamics Force Fields too Helical? *Biophys. J.* **2008**, *95*, L07−L09.

(3) Lange, O. F.; van der Spoel, D.; de Groot, B. L. Scrutinizing Molecular Mechanics Force Fields on the Submicrosecond Timescale with NMR Data. *Biophys. J.* **2010**, *99*, 647−655.

(4) Beauchamp, K. A.; Lin, Y.-S.; Das, R.; Pande, V. S. Are Protein Force Fields Getting Better? A Systematic Benchmark on 524 Diverse NMR Measurements. *J. Chem. Theory Comput.* **2012**, *8*, 1409−1414.

(5) Lindorff-Larsen, K.; Maragakis, P.; Piana, S.; Eastwood, M. P.; Dror, R. O.; Shaw, D. E. Systematic Validation of Protein Force Fields against Experimental Data. *PLOS One* **2012**, *7*, e32131.

(6) Cino, E. A.; Choy, W.-Y.; Karttunen, M. Comparison of Secondary Structure Formation Using 10 Different Force Fields in Microsecond Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2012**, *8*, 2725−2740.

(7) Rossetti, G.; Magistrato, A.; Pastore, A.; Carloni, P. Hydrogen Bonding Cooperativity in polyQ β-Sheets from First Principle Calculations. *J. Chem. Theory and Comput.* **2010**, *6*, 1777−1782.

(8) Fox, S. J.; Pittock, C.; Fox, T.; Tautermann, C. S.; Malcolm, N.; Skylaris, C.-K. Electrostatic embedding in large-scale first principles quantum mechanical calculations on biomolecules. *J. Chem. Phys.* **2011**, *135*, 224107.

(9) Salvador, P.; Kobko, N.; Wieczorek, R.; Dannenberg, J. J. Calculation of trans-Hydrogen-Bond $^{13}C-^{15}N$ Three-Bond and Other Scalar J-Couplings in Cooperative Peptide Models. A Density Functional Theory Study. *J. Am. Chem. Soc.* **2004**, *126*, 14190−14197.

(10) Beachy, M.; Chasman, D.; Murphy, R.; Halgren, T.; Friesner, R. Accurate ab Initio Quantum Chemical Determination of the Relative Energetics of Peptide Conformations and Assessment of Empirical Force Fields. *J. Am. Chem. Soc.* **1997**, *119*, 5908.

(11) DiStasio, R. A., Jr.; Steele, R. P.; Rhee, Y. M.; Shao, Y.; Head-Gordon, M. An improved algorithm for analytical gradient evaluation in resolution-of-the-identity second-order Møller-Plesset perturbation theory: Application to alanine tetrapeptide conformational analysis. *J. Comput. Chem.* **2007**, *28*, 839−856.

(12) Valdes, H.; Pluhackova, K.; Pitonak, M.; Rezac, J.; Hobza, P. Benchmark database on isolated small peptides containing an aromatic side chain: comparison between wave function and density functional theory methods and empirical force field. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2747−2757.

(13) Valdes, H.; Spiwok, V.; Rezac, J.; Reha, D.; Abo-Riziq, A. G.; de Vries, M. S.; Hobza, P. Potential-Energy and Free-Energy Surfaces of Glycyl-Phenylalanyl-Alanine (GFA) Tripeptide: Experiment and Theory. *Chem.—Eur. J.* **2008**, *14*, 4886−4898.

(14) Rossi, M.; Scheffler, M.; Blum, V. Impact of Vibrational Entropy on the Stability of Unsolvated Peptide Helices with Increasing Length. *J. Phys. Chem. B* **2013**, *117*, 5574−5584.

(15) Stearns, J. A.; Boyarkin, O. V.; Rizzo, T. R. Spectroscopic Signatures of Gas-Phase Helices: Ac-Phe-(Ala)5-Lys-H+ and Ac-Phe-(Ala)10-Lys-H+. *J. Am. Chem. Soc.* **2007**, *129*, 13820−13821.

(16) Stearns, J.; Seaiby, C.; Boyarkin, O.; Rizzo, T. Spectroscopy and conformational preferences of gas-phase helices. *Phys. Chem. Chem. Phys.* **2009**, *11*, 125.

(17) Xie, Y.; Schaefer, H., III; Dumitrescu, R.; Peng, B.; Li, Q.-s.; Stearns, J. A.; Rizzo, T. R. Conformational Preferences of Gas-Phase Helices: Experiment and Theory Struggle to Agree: The Seven-Residue Peptide Ac-Phe-(Ala)5-Lys-H+. *Chem.—Eur. J.* **2012**, *18*, 12941−12944.

(18) Wang, F. F.; Jenness, G.; Al-Saidi, W. A.; Jordan, K. D. Assessment of the performance of common density functional methods for describing the interaction energies of $(H_2O)_6$ clusters. *J. Chem. Phys.* **2010**, *132*, 134303−134303−8.

(19) Jenness, G. R.; Karalti, O.; Jordan, K. D. Benchmark calculations of water−acene interaction energies: Extrapolation to the water−graphene limit and assessment of dispersion-corrected DFT methods. *Phys. Chem. Chem. Phys.* **2010**, *12*, 6375.

(20) Santra, B.; Michaelides, A.; Scheffler, M. On the accuracy of density-functional theory exchange-correlation functionals for H bonds in small water clusters: Benchmarks approaching the complete basis set limit. *J. Chem. Phys.* **2007**, *127*, 184104.

(21) Santra, B.; Michaelides, A.; Fuchs, M.; Tkatchenko, A.; Filippi, C.; Scheffler, M. On the accuracy of density-functional theory exchange-correlation functionals for H bonds in small water clusters. II. The water hexamer and van der Waals interactions. *J. Chem. Phys.* **2008**, *129*, 194111.

(22) Kamariotis, A.; Boyarkin, O. V.; Mercier, S. R.; Beck, R. D.; Bush, M. F.; Williams, E. R.; Rizzo, T. R. Infrared Spectroscopy of Hydrated Amino Acids in the Gas Phase: Protonated and Lithiated Valine. *J. Am. Chem. Soc.* **2006**, *128*, 905−916.

(23) Gao, B.; Wyttenbach, T.; Bowers, M. T. Hydration of Protonated Aromatic Amino Acids: Phenylalanine, Tryptophan, and Tyrosineao. *J. Am. Chem. Soc.* **2009**, *131*, 4695−4701.

(24) Chutia, S.; Rossi, M.; Blum, V. Water Adsorption at Two Unsolvated Peptides with a Protonated Lysine Residue: From Self-Solvation to Solvation. *J. Phys. Chem. B* **2012**, *116*, 14788−14804.

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

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

(27) Adamo, C.; Barone, V. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *J. Chem. Phys.* **1999**, *110*, 6158.

(28) Tkatchenko, A.; Distasio, R. A.; Car, R.; Scheffler, M. Accurate and Efficient Method for Many-Body van der Waals Interactions. *Phys. Rev. Lett.* **2012**, *108*, 236402.

(29) Ambrosetti, A.; Reilly, A. M.; Robert A. DiStasio, J.; Tkatchenko, A. Long-range correlation energy calculated from coupled atomic response functions. 2013; manuscript to be submitted.

(30) Tkatchenko, A.; Ambrosetti, A.; Distasio, R. Interatomic methods for the dispersion energy derived from the adiabatic connection fluctuation-dissipation theorem. *J. Chem. Phys.* **2013**, *138*, 074106.

(31) DiStasio, R. A.; von Lilienfeld, O. A.; Tkatchenko, A. Collective many-body van der Waals interactions in molecular systems. *Proc. Nat. Acad. Sci.* **2012**, *109*, 14791−14795.

(32) Tkatchenko, A.; Rossi, M.; Blum, V.; Ireta, J.; Scheffler, M. Unraveling the Stability of Polypeptide Helices: Critical Role of van der Waals Interactions. *Phys. Rev. Lett.* **2011**, *106*, 118102.

(33) Rossi, M.; Blum, V.; Kupser, P.; von Helden, G.; Bierau, F.; Pagel, K.; Meijer, G.; Scheffler, M. Secondary Structure of Ac-Ala-LysH+ Polyalanine Peptides (n = 5,10,15) in Vacuo: Helical or Not? *J. Phys. Chem. Lett* **2010**, *1*, 3465−3470.

(34) Hua, S.; Xu, L.; Li, W.; Li, S. Cooperativity in Long α- and 3 10-Helical Polyalanines: Both Electrostatic and van der Waals Interactions Are Essential. *J. Phys. Chem. B* **2011**, *115*, 11462−11469.

(35) Kolb, B.; Thonhauser, T. Molecular Biology at the Quantum Level: Can Modern Density Functional Theory Forge the Path? *Nano LIFE* **2012**, *02*, 1230006.

(36) 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. *Comput. Phys. Commun.* **2009**, *180*, 2175.
(37) 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.
(38) Burke, K.; Ernzerhof, M.; Perdew, J. P. The adiabatic connection method: a non-empirical hybrid. *Chem. Phys. Lett.* **1997**, *265*, 115.
(39) Adamo, C.; Cossi, M.; Scalmani, G.; Barone, V. Accurate static polarizabilities by density functional theory: Assessment of the PBE0 model. *Chem. Phys. Lett.* **1999**, *307*, 265−271.
(40) *TINKER - Software Tools for Molecular Design*, version 5.1; Jay Ponder Lab, Washington University: St Louis, MO, http://dasher.wustl.edu/tinker/, 2013.
(41) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins: Struct., Funct. Bioinf.* **2006**, *65*, 712−725.
(42) Jorgensen, W.; Maxwell, D.; Tirado-Rives, J. Development and testing of the OPLS all-atom force field on the conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225.
(43) Karplus, M.; Kolker, H. J. Van der Waals Forces in Atoms and Molecules. *J. Chem. Phys.* **1964**, *41*, 3955−3961.
(44) Fabri, C.; Szidarovszky, T.; Magyarfalvi, G.; Tarczay, G. Gas-Phase and Ar-Matrix SQM Scaling Factors for Various DFT Functionals with Basis Sets Including Polarization and Diffuse Functions. *J. Phys. Chem. A* **2011**, *115*, 4640−4649.
(45) Merrick, J. P.; Moran, D.; Radom, L. An Evaluation of Harmonic Vibrational Frequency Scale Factors. *J. Phys. Chem. A* **2007**, *111*, 11683−11700.
(46) Gloaguen, E.; de Courcy, B.; Piquemal, J.-P.; Pilmé, J.; Parisel, O.; Pollet, R.; Biswal, H. S.; Piuzzi, F.; Tardivel, B.; Broquier, M.; et al. Gas-Phase Folding of a Two-Residue Model Peptide Chain: On the Importance of an Interplay between Experiment and Theory. *J. Am. Chem. Soc.* **2010**, *132*, 11860−11863.
(47) Heyd, J.; Scuseria, G. E.; Ernzerhof, M. Hybrid functionals based on a screened Coulomb potential. *J. Chem. Phys.* **2003**, *118*, 8207.
(48) Krukau, A. V.; Vydrov, O. A.; Izmaylov, A. F.; Scuseria, G. E. Influence of the exchange screening parameter on the performance of screened hybrid functionals. *J. Chem. Phys.* **2006**, *125*, 224106.
(49) Levchenko, S.; Ren, X.; Wieferink, J.; Rinke, P.; Blum, V.; Scheffler, M.; Johanni, R. Unpublished.
(50) Ochsenfeld, C.; White, C. A.; Head-Gordon, M. Linear and sublinear scaling formation of Hartree-Fock-type exchange matrices. *J. Chem. Phys.* **1998**, *109*, 1663.