JCTC Journal of Chemical Theory and Computation

Article

Cite This: J. Chem. Theory Comput. 2018, 14, 5406−5418

pubs.acs.org/JCTC

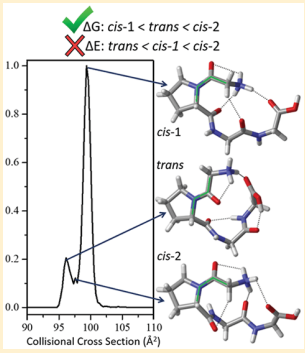
Electronic Energies Are Not Enough: An Ion Mobility-Aided, Quantum Chemical Benchmark Analysis of H^+GPGG Conformers

Daniel Beckett, Tarick J. El-Baba, David E. Clemmer, and Krishnan Raghavachari*

Department of Chemistry, Indiana University, Bloomington, Indiana 47405, United States

Supporting Information

ABSTRACT: Peptides and proteins exist not as single, lowest energy structures but as ensembles of states separated by small barriers. In order to study these species we must be able to correctly identify their gas phase conformational distributions, and ion mobility spectrometry (IMS) has arisen as an experimental method for assessing the gas phase energetics of flexible peptides. Here, we present a thorough exploration and benchmarking of the low energy conformers of the small, hairpin peptide H^+GPGG with the aid of ion mobility spectrometry against a wide swath of density functionals (35 dispersion-corrected and uncorrected functionals represented by 21 unique exchange−correlation functionals) and wave function theory methods (15 total levels of theory). The three experimentally resolved IMS peaks were found to correspond to three distinct pairs of conformers, each pair composed of species differing only by the chair or boat configuration of the proline. Two of the H^+GPGG conformer pairs possess a *cis* configuration about the Pro−Gly[1] peptide bond while the other adopts a *trans* configuration. While the experimental spectrum reports a higher intensity for the *cis*-1 conformer than the *trans* conformer, 13 WFT levels of theory, including a complete basis set CCSD(T) extrapolation, obtain *trans* to be favored in terms of the electronic energy. This same effect is seen in 14 of the 18 dispersion-corrected density functionals studied, whereas the remaining 17 functionals show more variety. Only when Gibbs free energies are considered do the WFT methods and dispersion-corrected functionals reflect the experimental distribution. CAM-B3LYP-D3BJ emerges as the best-performing density functional, matching the experimental distribution and the CCSD(T)/CBS relative energies within 6%. Further analysis reveals the *trans* conformer to be favored electronically, but entropically disfavored, leading to the experimental preference of the *cis*-1 conformer. These results highlight the danger in considering only the electronic energies, which is common practice in electronic structure theory predictions of conformational energy distributions. Additionally, the effects of temperature and scaling of the frequencies used in obtaining the Gibbs free energies are explored.



1. INTRODUCTION

In order to predict the biological activity of a polypeptide it is essential to understand, not just the native lowest energy geometry, but the dynamics and the accessible transient structures of the biomolecule in question.[1] Accordingly, conformationally flexible molecules exist as an ensemble of states, some of which may possess fundamentally different structural motifs and correspondingly dissimilar functions. Currently, an increasingly popular method for understanding the conformational landscape of a polypeptide is ion mobility spectrometry (IMS) coupled to mass spectrometry (MS). While MS has been used to determine the stoichiometry of protein−protein and protein−ligand complexes,[2] the tandem IMS-MS technique builds off this information by allowing the separation of structures based primarily on their shapes. Due to this complementary information, IMS-MS studies have assisted in the elucidation of previously unobserved biomolecular interactions.[2−6]

In IMS, ions migrate through a drift tube filled with a buffer gas under the influence of a weak electric field. The unique mobility of an ion is determined by the time it takes, t_D, to reach the detector at the end of the drift tube as it travels through an inert (typically N_2 or He) buffer gas. An ion with a larger rotationally averaged collisional cross section will collide with the buffer gas more often than an ion with a smaller collisional cross section, resulting in a longer drift time.[7] From the parameters of the IMS experiment, it is possible through eq 1 to derive the collisional cross section, Ω, which is a measure of an ion's shape.[8] Here, ze is the charge of the ion, m_I the mass of the ion, m_B the mass of the buffer gas, N the number density of the buffer gas, E the electric field strength, and L the drift tube length.

$$\Omega = \frac{(18\pi)^{1/2}}{16}\frac{ze}{(k_B T)^{1/2}}\left[\frac{1}{m_I}+\frac{1}{m_B}\right]^{1/2}\frac{t_D E}{L}\frac{760}{P}\frac{273}{T}\frac{1}{N} \quad (1)$$

Though the cross section is certainly an invaluable quantity, the actual three-dimensional structure it represents must be found and confirmed using computational chemistry. Fortunately, several methods exist that calculate collisional cross sections from trial geometries.[9,10] In this work we use the trajectory method developed by Schatz and co-workers which includes a Lennard-Jones potential to approximate the long-

Received: June 26, 2018
Published: September 7, 2018

DOI: 10.1021/acs.jctc.8b00648
J. Chem. Theory Comput. 2018, 14, 5406−5418

range interactions between the ion and the buffer gas, as opposed to a simple hard sphere projection method.[11]

The ability to match experimental collisional cross sections to calculated structures facilitates a direct route for the confirmation of global, ground state geometries and is therefore a very attractive prospect in computational chemistry. Matching the collisional cross sections of small, conformationally flexible peptides with their gas phase geometries at multiple levels of theory will provide the groundwork for high throughput quantum chemical procedures designed to process large peptides with confidence. Additionally, it has been shown previously that taking advantage of the motifs present in polypeptides can be a key component required to obtain accurate results from fragmentation-based methods.[12,13] By this logic, a careful and fundamental exploration of the conformational landscapes of small peptides is warranted to better understand motifs within larger structures and to identify the means by which to confirm them.

Gly-Pro-Gly-Gly (GPGG) represents an interesting species as it is among the simplest hairpin peptides and possesses an implicit structural heterogeneity in the two possible *cis* or *trans* arrangements of the N-terminal Gly and Pro alpha carbons. The collisional cross section distribution of singly protonated Gly-Pro-Gly-Gly (H^+GPGG) has been reported previously by the Clemmer and Rizzo groups.[14,15] While previous work confirmed the *cis* conformer as the gas phase lowest energy structure, multiple potential structures were presented within 1.5 kcal/mol of each other with differing intramolecular hydrogen-bonding skeletons.[14] The current work aims to perform a thorough analysis of H^+GPGG in an effort to determine an accurate and high throughput protocol for the analysis of similar conformationally flexible peptides with complex intramolecular hydrogen bonds.

An in-depth study of the six lowest-lying H^+GPGG conformers is reported, testing a range of density functionals against high level gas phase calculations to determine which set of approximations best match both experiment and high-level calculations. A range of approximations to the MP2 and CCSD(T) energies are discussed and evaluated with respect to their ability to predict CCSD(T)/CBS conformational energies. The zero-point corrected energies and Gibbs free energies are discussed to determine which should be used as the comparison to experimental intensities. Non-negligible entropic effects are found to drive the experimental distribution, and previous successful assignments are shown to be based on incomplete models due to fortuitous cancellation of errors. The robustness of the relative Gibbs free energies is checked through the use of frequency scaling factors and found to be unaffected. The effect of temperature on the calculated Gibbs free energy distribution is also examined to obtain an approximate working temperature of molecules in the IMS drift tube.

2. METHODS

2.1. Conformational Analysis. To obtain initial conformers, the PCMODEL conformational analysis software[16] was used to build H^+GPGG in a β-strand configuration. Conformers were obtained by stochastically choosing a set of bonds to rotate, quenching the result using the MMFF94 force field[17] and discarding structures outside a 7 kcal/mol energy window relative to the lowest energy conformer. Generated conformers were optimized with the PM6 semi-empirical method[18] as implemented in the Gaussian Development Version (GDV v. I06) computational chemistry package.[19] Degenerate structures from the PM6 analysis were discarded and the geometries were further optimized, and frequencies obtained, at the DFT/6-311++G(d,p) level of theory, where DFT corresponds to a list of selected density functional approximations. For each geometry the collisional cross section was determined using the trajectory method as implemented in MOBCAL.[11] For each MOBCAL calculation, the same starting conditions were used to allow for unbiased comparisons between different levels of theory. This study reports the six lowest energy conformers across all examined levels of theory, though it should be noted that many other, higher energy conformers were also obtained and may contribute to minor details in the experimental spectrum (such as peak widths and slight shoulders). The six conformers discussed here are enough to probe the three major peaks in the spectrum in terms of collisional cross section and relative intensity.

To compare the results of theory with the experimental spectrum, a simple Boltzmann distribution is invoked (eq 2) wherein T corresponds to temperature (298.15 K unless specified), ΔE to the difference in conformer energies, and k_B to the Boltzmann constant. In all calculations in the text, ΔE is calculated with respect to the conformer referred to as *cis*-1, as it was found to make up the major peak in the experimental spectrum. Additionally, all peaks in the experimental spectrum are normalized with respect to the *cis*-1 peak, and we accordingly normalize all Boltzmann probabilities with respect to this peak.

$$p_{\text{norm}} = e^{-\Delta E/k_B T} \tag{2}$$

2.2. Density Functionals. A number of density functionals were tested to determine the best fit with experiment and high-level calculations. The GGA (generalized gradient approximation) double-hybrid B2PLYP functional including MP2 correlation,[20] the reparameterized "general purpose" double-hybrid B2GPPLYP,[21] the popular B3LYP hybrid GGA functional,[22−25] the nonempirical GGA hybrid PBE0,[26] the one-parameter GGA hybrid mPW1PW91,[27] the one-parameter meta-GGA hybrid TPSSh,[28] the local meta-GGA functional TPSSTPSS,[29] and the revPBE GGA functional[30,31] were tested as the main "uncorrected" density functionals in this study. Additionally, we probed a wide swath of the Minnesota meta-GGA functionals: the M06-L local meta-GGA,[32] the M06-2X meta-GGA hybrid,[33] the M06-HF hybrid with full Hartree−Fock exchange,[34] the local meta-NGA (nonseparable gradient approximation) MN15-L,[35] and the meta-NGA hybrid MN15.[36]

On top of the density functionals described above, we also explored the effects of long-range and dispersion corrections on the conformational energy distribution. The long-range corrected functionals explored include: CAM-B3LYP,[37] LC-ωPBE,[38] ωB97X-D with built-in D2 dispersion corrections,[39] ωB97X-D3 with built in D3(0) corrections,[40] and X3LYP.[41] These range-separated, hybrid functionals shift the amount of Hartree−Fock exchange from a "short-range" to a larger "long-range" value, differing on the values of these parameters and the function varying the percentage of exact exchange. To understand the effect of dispersion, we added Grimme's D3 correction[42] with Becke−Johnson damping[43] to all studied functionals for which parameters were readily available, forming: B2PLYP-D3BJ,[44] B2GPPLYP-D3BJ,[44] B3LYP-D3BJ,[43] CAM-B3LYP-D3BJ,[44] X3LYP-D3BJ,[45] PBE0-D3BJ,[43] LC-ωPBE-D3BJ,[44] TPSSh-D3BJ,[44] TPSSTPSS-D3BJ,[43] and revPBE-D3BJ.[43] Additionally, we examined the hybrid GGA B97-D3BJ functional due to its applications in biomolecular

DOI: 10.1021/acs.jctc.8b00648
J. Chem. Theory Comput. 2018, 14, 5406−5418

systems,[46] as well as the dispersion-corrected GGA double-hybrid DSD-PBEP86-D3BJ (as the functional was parameterized for use with dispersion corrections, the uncorrected form is left out; DSD-PBEP86-D3BJ is often shortened to DSD-PBEP86 in many computational programs, including Gaussian, but we include the suffix for clarity and note that the SCS and D3BJ parameters come from the revised parameterization, not the original communication).[47,48] Finally, we examined the zero-damped forms of the D3 correction for the Minnesota functionals for which parameters were available: M06-L-D3(0), M06-2X-D3(0), and M06-HF-D3(0).[44] The majority of density functional calculations were done in the Gaussian Development Version (GDV v. I06) computational chemistry package using an ultrafine grid and tight SCF convergence criteria.[19] DSD-PBEP86-D3BJ calculations were performed with Gaussian 16 (G16 v. A03).[49] Version 4.0.0.2 of the ORCA computational chemistry program was used for the revPBE, revPBE-D3BJ, and ωB97X-D3 calculations with the Grid6 integration grid, Grid7 for the final energy evaluation, and tight SCF convergence.[50]

2.3. Wave Function Theory Methods. The optimized B3LYP geometries of the six lowest in energy conformers were used as starting points for MP2/6-311++G(d,p) geometry optimization and frequency calculations.[51−53] All wave function theory method applications in this analysis correspond to single-point energy calculations on the optimized MP2 geometry with associated MP2 frequencies and thermochemical quantities. All thermochemical quantities were obtained within the rigid rotor harmonic oscillator (RRHO) approximation at 298.15 K and a pressure of 1 atm unless otherwise noted.

RI-MP2/jul-cc-pVXZ calculations[54,55] (with the corresponding correlation-fitted aug-cc-pVXZ/C auxiliary basis sets for the resolution of identity, RI, approximation) were performed using version 4.0.0.2 of the ORCA computational chemistry program, where X corresponds to double (D), triple (T), and quadruple (Q) zeta basis sets. Additionally, MP2/jul-cc-pVDZ single-point calculations were performed with Gaussian to confirm the accuracy of the RI approximation. The jul-cc-pVXZ basis sets follow the convention of Truhlar and refer to the appropriate aug-cc-pVXZ basis set minus the diffuse functions on hydrogen.[56] RI-MP2/aug-cc-pVQZ single-point calculations were undertaken to assess the effect of the hydrogen diffuse functions.[57,58]

To reach the theoretical high level in this study, CCSD(T)/jul-cc-pVDZ single-points were obtained using version 1.1 of the Psi4 quantum chemistry program. The CCSD(T) complete basis set limit (CBS) was approximated through the additivity approach reported by Jurecka and Hobza as shown in eq 3, and this functions as the theoretical high level in our study.[59] Here, RI-MP2/CBS refers to the extrapolated RI-MP2 value following the global two point extrapolation forwarded by Halkier et al. as shown in eq 4, and J′VXZ refers to the appropriate jul-cc-pVXZ basis set.[60] For the RI-MP2/CBS extrapolation (referred to simply as MP2/CBS in the text), X corresponds to 3 (the jul-cc-pVTZ basis) and Y corresponds to 4 (the jul-cc-pVQZ basis). The value of α in eq 3 was set to 3 for the MP2 correlation component and 5 for the HF reference.[60] The effect of leaving α equal to 3 for both components of the MP2 energy was also explored, however the CCSD(T)/CBS obtained with an MP2/CBS with α = 3, 5 will be used as the main computational reference for the analysis to follow (vide infra).

$$E(\mathrm{CCSD(T)/CBS}) \approx E(\mathrm{RI\text{-}MP2/CBS}) + E(\mathrm{CCSD(T)/J'VDZ}) - E(\mathrm{RI\text{-}MP2/J'VDZ}) \quad (3)$$

$$E(\mathrm{CBS}) = \frac{E(\mathrm{VXZ})X^{\alpha} - E(\mathrm{VYZ})Y^{\alpha}}{X^{\alpha} - Y^{\alpha}} \quad (4)$$

The effect of the basis set on the perturbative triples correction was investigated by adding the triples component of CCSD(T)/cc-pVDZ single-points to the CCSD energy of jul-cc-pVDZ calculations, this approach is referred to as CCSD/jul+(T)/cc-pVDZ. Additionally, to determine the performance of linear-scaling approximations to correlated methods, single-point DLPNO-CCSD(T)/jul-cc-pVXZ calculations[61,62] (with aug-cc-pVXZ/C auxiliary basis sets) were performed using version 4.0.0.2 of ORCA computational chemistry program[50] with tight DLPNO parameters (for more details on DLPNO parameters and the rationale for the use of tight parameters, the interested reader may consult section 3 of the Supporting Information).[63]

2.4. Frequency Scaling. All thermochemical quantities were obtained without scaling factors due to the near unity value of the scaling factor for the majority of the density functionals considered and the lack of predefined scaling factors for every tested functional, potentially leading to an imbalance in the analysis. However, MP2 harmonic oscillator frequencies are known to, in general, be overestimated by about 4% across most tested basis sets.[64] As the frequencies are only being used in this case to determine thermochemical properties and relative thermochemical properties at that, we expect the effect of scaling to be negligible and report all wave function theory methods with unscaled MP2 properties. However, to the best of our knowledge, there has been little investigation into the effect of frequency scaling on relative Gibbs free energies for conformers, and we have used this as an opportunity to explore and expand upon the application of scaling factors to these quantities.

The Supporting Information contains a detailed description of the effects of frequency scaling factors upon the conformational energy distribution. In general, when treated carefully, we found the effect to be negligible, and we do not include these results in the main text of this study. Aside from considering one single scaling factor, we investigated the effect of using two scaling factors: one for "low" frequencies and the other for "high" frequencies. This method, of course, relies heavily upon the cutoff value distinguishing low frequencies from high frequencies. In the Supporting Information we present a novel method for attaining this value and demonstrate the danger of carelessly choosing this cutoff value to be a single, constant frequency when the treated conformers have differing numbers of frequencies below the cutoff. However, even when using a cutoff value far outside the ideal value, the relative Gibbs free energies vary by less than 0.1 kcal/mol. We plan to investigate systems where intelligent choice of a cutoff value makes a large difference in experimental comparison, but for now we find any reasonable choice of scaling factors or cutoff values makes little difference to the conformational distribution. We encourage the interested reader to peruse section 1 of the Supporting Information.

2.5. Procedure for Experimental Spectrum. GPGG was purchased from Sigma-Aldrich (St. Louis, MO) and prepared for electrospray ionization at ∼10 μM in 50:50 water:methanol.

DOI: 10.1021/acs.jctc.8b00648
J. Chem. Theory Comput. 2018, 14, 5406−5418

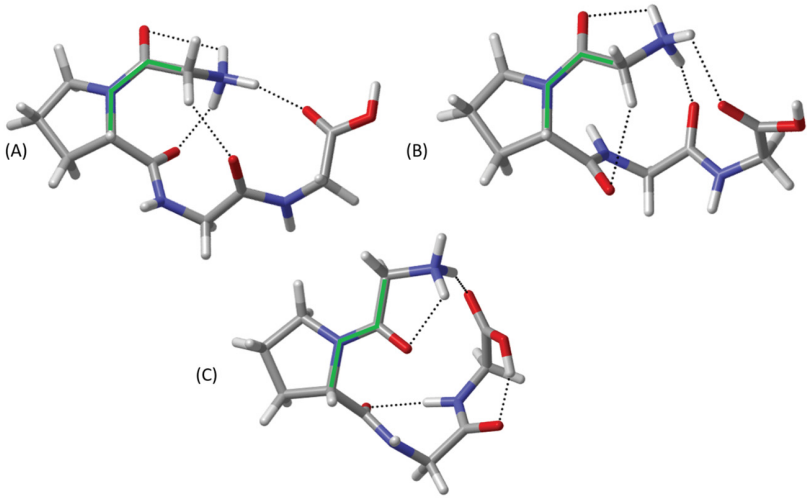


Figure 1. CAM-B3LYP-D3BJ/6-311++G(d,p) optimized geometries of the three lowest energy H^+GPGG conformers, prolines in the chair conformation relative to the C-terminus. A and B are referred to in the text as *cis*-c1 (or *cis*-1) and *cis*-c2 (or *cis*-2) while C is the lowest energy *trans* structure, referred to as *trans*-c (or *trans*). Dotted lines drawn for nonbonded hydrogen to heavy atom distances less than 2.6 Å. Green line drawn to illustrate *cis* vs *trans* geometries of the proline α carbon and N-terminal α carbon.

Details of IMS theory[8−11] and instrumentation[65] are provided elsewhere; only a brief description of the instrument is given here. Ions are produced by electrospray ionization using a Triversa Nanomate autosampler (Advion, Ithica, NY) and stored in an hourglass-shaped ion funnel.[66] Using an electrostatic ion gate, ion packets are periodically released into the 3-m drift tube filled with 3.00 ± 0.03 Torr He buffer gas, held at ∼10 V cm^{-1}. Ions are separated in the drift tube based on differences in their unique mobilities, which are related to their sizes and shapes. Mobility separated ions eluting from the drift tube are pulsed into an orthogonal time-of-flight mass spectrometer for analysis of their mass-to-charge (m/z) ratios. The focus of this work is the accurate quantum chemical calculations of the gas phase H^+GPGG conformer distributions. Thus, ion storage conditions were optimized to reflect the gas phase quasi-equilibrium distributions.[67]

3. RESULTS AND DISCUSSION

3.1. H^+GPGG Low Energy Conformers. Figure 1 shows the three lowest in energy H^+GPGG conformers throughout all levels of theory with the proline in the "chair" configuration. For each conformer in Figure 1, there is a closely related conformer where the proline ring adopts the "boat" configuration with respect to the C-terminal residues (vide infra). The green line in Figure 1 follows the proline α carbon to the N-terminal glycine α carbon to illustrate *cis* vs *trans* structures. The conformation in Figure 1A is referred to as *cis*-c1, "c" indicating the chair conformation of the proline group, and has previously been found to be the lowest energy structure at the B3LYP/6-311++G(d,p) level of theory.[14] Figure 1B illustrates structure *cis*-c2, which differs from *cis*-c1 by a rotation of the bond between the alpha and carbonyl carbons of proline. The rotation in *cis*-c2 switches the hydrogen bonds engaged in by the Pro and Gly^3 carbonyl groups. For convenience, these two *cis* conformations will be referred to as *cis*-1 and *cis*-2 in the rest of the paper, unless specific relative values are under discussion.

All assigned hydrogen bonds in *cis*-1 and *cis*-2, except for the CH···O interactions, have been corroborated by IR analysis in the previous study.[14] Our calculations show strong evidence for a CH···O interaction; taking the CAM-B3LYP-D3BJ parameters (vide infra), the H to O distance in *cis*-c1 is 2.3 Å, falling within the weak category, and the bond angle is 143°, within the moderate range limit of over 130°.[68] In contrast, the CH···O interaction in *cis*-c2 has a length of 2.6 Å and a moderate bond angle of 148°. It is important to note that such weak and moderate hydrogen bonds are dispersion-dominated, meaning any method neglecting the effects of dispersion is likely to treat these conformers incorrectly, and the importance of dispersion corrections in predicting peptide structures has been previously demonstrated in both the gas and crystalline phases.[69,70]

The conformer shown in Figure 1C is the chair conformation of the only low-lying *trans* structure, referred to as *trans*-c (or *trans*). The *trans* specimen is stabilized by an intricate hydrogen bonding network consisting of four hydrogen bonds within the moderate strength range of 1.5−2.2 Å, two of which (the two C-terminal hydrogen bonds) are on the order of 1.6 Å across all levels of theory. On the other hand, the two *cis* species are each stabilized by two hydrogen bonds in the moderate category along with two others spanning distances within the weak range (above 2.2 Å).

Finally, we consider both the boat and chair forms of each conformer throughout these analyses. Here we use "chair" to describe conformations wherein the proline γ carbon is oriented opposite of the Gly-Gly C-terminal strand with respect to the plane defined by the proline amine nitrogen, α carbon, and δ carbon, and "boat" corresponds to the conformation wherein the γ carbon lies on the same side of the plane as the Gly-Gly strand. The "chair" and "boat" distinction is simply a nomenclatural convenience, and these conformers have similar dihedral angles as the exo and endo envelope structures from the proline puckering literature.[71−73] The chair/boat energy difference is usually small, and as they share the same network of hydrogen bonds, the barrier to interconvert between them is also relatively low however both orientations exist as discrete minima. In all sections to follow we treat the intensities as derived from the lowest energy member of each conformer pair. While experimental intensities may be slightly affected by the inclusion of both chair and boat species, we believe that reporting the lowest energy conformer of each pair for all three observed peaks minimizes the error in the overall analysis.

DOI: 10.1021/acs.jctc.8b00648
J. Chem. Theory Comput. 2018, 14, 5406−5418

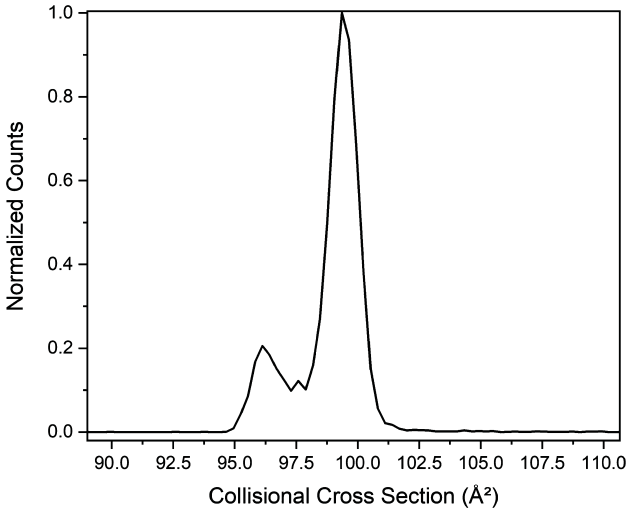


Figure 2. Experimental ion mobility collisional cross section distribution of H^+GPGG.

3.2. Collisional Cross Sections. Figure 2 shows the experimental collisional cross section distribution we will use as reference throughout this text. The experimental spectrum consists of a broad, low intensity peak (0.21 normalized counts) at 96.1 Å^2, a sharper high intensity peak (1 normalized count, the reference peak for all normalizations in this study) at 99.3 Å^2, and a low intensity feature (0.12 normalized counts) in between them centered at 97.6 Å^2. The same trends are evident across all calculated cross sections (Figure S1 of the Supporting Information): the *cis*-1 species have the highest cross sections, followed by *cis*-2 and *trans*. This agrees with the previous work which assigned the low Å^2 peak to *trans* and the higher Å^2 peak to *cis*-1.[14] The sharp peak in-between the *trans* and *cis*-1 peaks has not been previously reported and fits well with the *cis*-2 cross sections.

To be sure of these assignments, we consider the MP2 collisional cross sections versus experiment. The *cis*-c1 collisional cross section of 99.6 Å^2 matches very well with the large peak at 99.3 Å^2 and the *cis*-c2 collisional cross section of 98.0 Å^2 matches equally well with the diminutive peak at 97.6 Å^2. Finally, the experimental peak with the smallest cross section (96.1 Å^2)

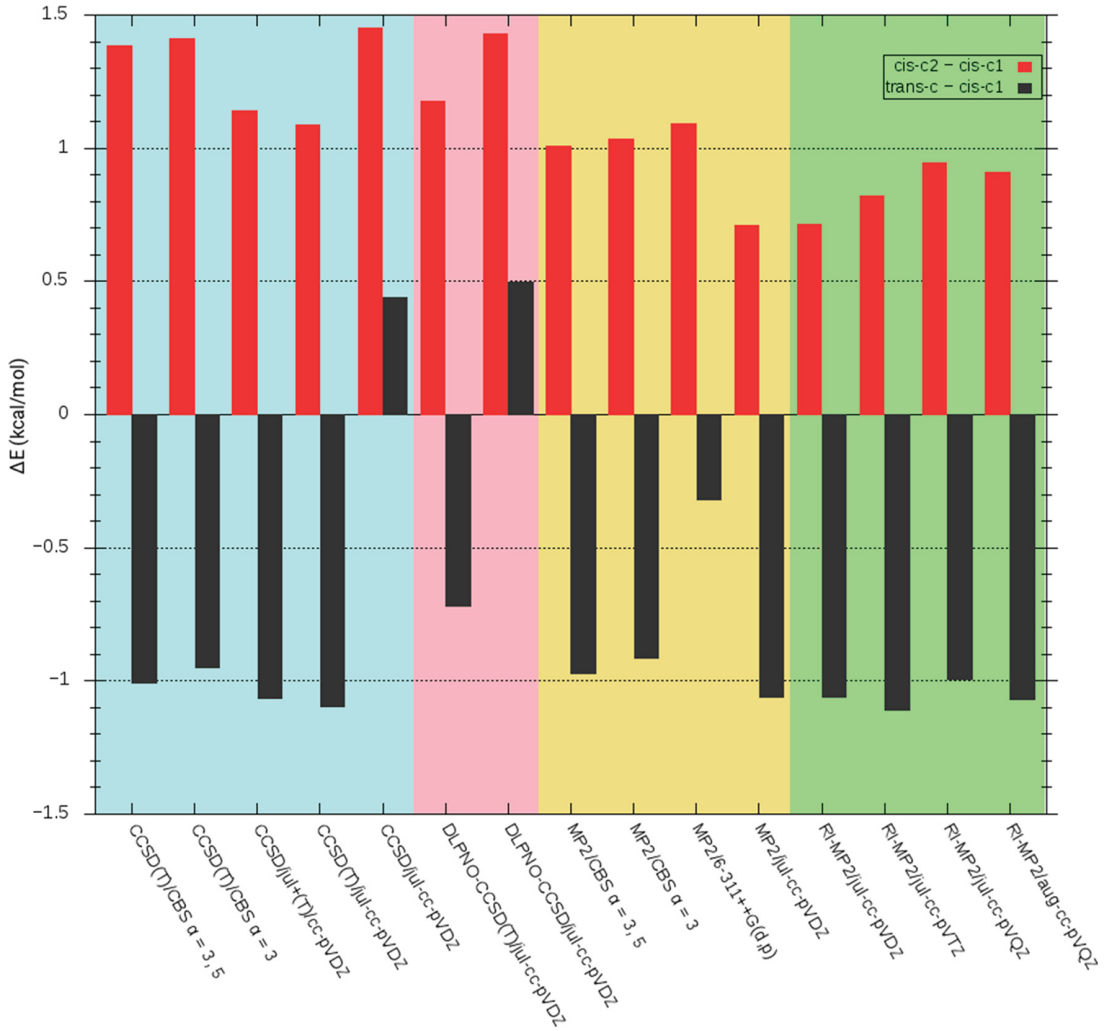


Figure 3. Post-Hartree−Fock wave function theory electronic energies of H^+GPGG conformers relative to the *cis*-c1 energy. Red bars represent *cis*-c2 relative energies, black bars represent *trans*-c relative energies. CBS values obtained as described in the Methods section. DLPNO refers to DLPNO with TightPNO cutoffs.

DOI: 10.1021/acs.jctc.8b00648
J. Chem. Theory Comput. 2018, 14, 5406−5418

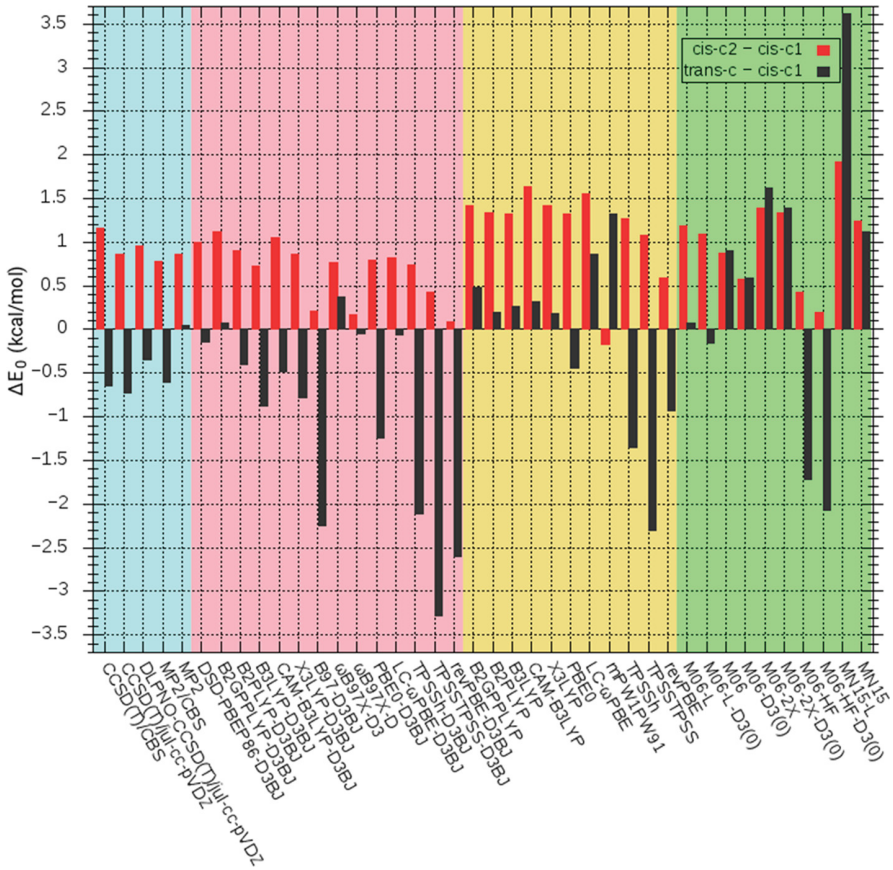


Figure 4. Zero-point corrected energies of H^+GPGG conformers relative to *cis*-c1. Red bars represent *cis*-c2 relative energies, black bars represent *trans*-c relative energies. CBS values obtained as described in the Methods section. All geometries and vibrational corrections for WFT methods (light blue section) obtained with MP2/6-311++G(d,p), all DFT properties obtained with DFT/6-311++G(d,p). DLPNO with TightPNO cutoffs.

corresponds to the MP2 *trans*-c CCS value of 94.9 Å^2. In this manner we assign the peaks in the experimental spectrum to correspond to, from left to right: *trans*, *cis*-2, *cis*-1, from MP2/6-311++G(d,p) geometries. Section 2 of the Supporting Information provides an analysis of the DFT collisional cross sections in comparison to experiment and the MP2 values reported here. Every examined density functional matches the MP2 assignments of the experimental peaks, with no collisional cross section deviating from the experimental values by more than 3%, the suggested error margin for MOBCAL.[74] All calculated collisional cross sections can be found in Table S8 of the Supporting Information.

3.3. Wave Function Theory, Electronic Energies. To begin analyzing the computational predictions of the experimental peak intensities displayed in Figure 2, Figure 3 presents the post-Hartree−Fock, wave function theory, electronic energies of the chair conformers relative to *cis*-c1. If the zero-point energy and other effects are assumed to be similar for the different conformers, the electronic energies should be sufficient to understand their experimental peak intensities. However, this is not the case as shown by the observation that, with the exception of CCSD, all levels, including CCSD(T) which should be an improvement over CCSD, place *trans* lower in energy than *cis*-1. These results are inconsistent with the experimental peak intensity of the *trans* conformer which is only 21% of the *cis*-1 intensity. Clearly, the electronic energies alone are insufficient to understand the peak intensities in the experiment. Taking this observation into consideration, we abandon experimental comparison in this section and instead focus our attentions on the computational high level, CCSD(T)/CBS, to understand which levels of theory best replicate the most trusted level of theory.

CCSD(T)/CBS (α = 3, 5) was obtained as described in eqs 3 and 4 of the Methods section with the lower α value approximating the slow convergence of the MP2 correlation energy and the higher value corresponding to the much faster convergence of the Hartree−Fock reference wave function. CCSD(T)/CBS (α = 3, 5) predicts *trans*-c to be 1.01 kcal/mol more favorable than *cis*-c1, and *cis*-c2 to be 1.38 kcal/mol less favorable than *cis*-c1. While experimental comparison is not useful at this level, there are some comparisons to be made to assess the validity of the CBS extrapolation. First, the CCSD(T)/CBS values with α = 3 for the entire MP2/CBS energy are quite close to the split α values: differences of 0.06 and 0.03 kcal/mol for the *trans*-c and *cis*-c2 relative energies, respectively. This small difference ensures that the effect of α in eq 4 is very mild, which is desirable for a potentially adjustable parameter (for example, a recent study on the accuracy of MP2-F12 calculations on small peptides employed an optimized α of 3.09).[75] The fourth pair of columns in Figure 3 reports the CCSD(T)/jul-cc-pVDZ energies, the most expensive calculations undertaken in this work. The jul-cc-pVDZ relative energies are only 0.16 kcal/mol higher and lower than the CCSD(T)/CBS results for *cis*-c2 and *trans*-c, respectively.

The closeness of the double-ζ and CBS values, as well as the closeness of the MP2 and CCSD(T) jul-cc-pVDZ results as they

DOI: 10.1021/acs.jctc.8b00648
J. Chem. Theory Comput. 2018, 14, 5406−5418

relate in eq 3, further cements our confidence in the mild nature of the CBS approximation as applied to relative conformer energies. As the basis set increases, the MP2 results more closely resemble the CCSD(T) values and we also notice that the difference between MP2/jul-cc-pVQZ and MP2/aug-cc-pVQZ results is quite small (the largest difference being less than 0.08 kcal/mol), indicating the hydrogen diffuse functions in the aug basis set do not play a large role in the conformer energies. Additionally, in the interest of finding accurate extrapolations able to be applied to slightly larger molecules, the third pair of bars in Figure 3 presents the relative energies of the CCSD/jul-cc-pVDZ calculations with the costly perturbative triples contribution, (T), computed with the smaller cc-pVDZ basis set. This approximation, dubbed CCSD/jul+(T)/cc-pVDZ, produces relative energies less than 0.1 kcal/mol off from the full CCSD(T)/jul-cc-pVDZ values for each conformer. The final pair of bars in the blue section of Figure 3 presents the CCSD relative energies and we see the triples correction is clearly the main force favoring the *trans* conformers over the *cis*. The CCSD/jul-cc-pVDZ relative *trans* energy is 1.60 kcal/mol higher than with CCSD(T)/jul-cc-pVDZ, and 0.44 kcal/mol higher than *cis*-1. In terms of raw comparison with experiment, the CCSD values qualitatively match the relative intensities, but the *trans* intensity is much higher (47.5% normalized population compared to an experimental value of 21%) than the anticipated value.

In addition to the extrapolations to the CCSD(T) energy presented above, two methods that reduce the scaling of costly correlation energy calculations are analyzed: RI-MP2 and DLPNO-CCSD(T). RI-MP2 (resolution of identity MP2) significantly reduces the scaling of MP2 by approximating the four center integrals as three center integrals with an auxiliary basis set expansion.[54] As can be seen in Figure 3, the greatest deviation between the MP2/jul-cc-pVDZ and RI-MP2/jul-cc-pVDZ relative energies is 0.004 kcal/mol, marking the RI approximation as an incredibly safe method for approximating the MP2 energies and obtaining the MP2/CBS values used in the CCSD(T) extrapolations. DLPNO-CCSD(T) (domain-based local pair-natural orbital CCSD(T)) is a linear scaling approximation to the CCSD(T) energy which achieves impressive speed up by employing extensive screening. The DLPNO approximation employs a localized RI-MP2-based method to determine which pair-natural orbitals (PNOs, in the case of CCSD; triple-natural orbitals, TNOs, used in the perturbative triples component) should be included in the final CCSD(T) calculations.[61,62] As with any method employing numerical cutoffs, however, this leaves DLPNO-CCSD(T) susceptible to the values of these cutoffs, which are optimized to achieve a balance between speed and accuracy.

Due to poor behavior exhibited by the default, NormalPNO, cutoffs (the reader interested in a more thorough analysis of DLPNO values is directed to section 3 of the Supporting Information), we employ TightPNO cutoff values here in the pink section of Figure 3 and throughout the paper.[63] For the *cis*-c2 relative energies, TightPNO performs remarkably well, with the largest difference from the unaccelerated CCSD and CCSD(T) energies being 0.09 kcal/mol. However, the *trans*-c relative energy with DLPNO−CCSD(T) is 0.38 kcal/mol higher than the CCSD(T) value, indicating an issue with the cutoffs employed for the triples correction for this particular species. Finally, it should be noted that only "chair" conformers are considered in Figure 3. A description of the chair vs boat distribution is available in section 4 of the Supporting Information; however, at this level of theory the difference is negligible, and all relative energies can be found in Table S2.

3.4. Zero-Point Corrected Energies and DFT. Upon considering the zero-point corrected energies presented in Figure 4, the first *cis*-1/*trans* orderings consistent with experiment appear. Interestingly, of the 40 methods included in Figure 4, 16 correctly predict *trans* and *cis*-2 to be less favorable than *cis*-1, but only 12 methods also favor *trans* compared to *cis*-2. Every dispersion-corrected functional (aside from the dispersion-corrected Minnesota functionals, ωB97X-D3, and B2GPPLYP-D3BJ where the *cis*-1 and *trans* species are separated by less than 0.05 kcal/mol), reported in pink, favors *trans* over *cis*-1, as well as every WFT method aside from MP2/6-311++G(d,p). Given the basis set convergence of the MP2 results approaching the CCSD(T) limit, the "correct" ordering predicted by MP2 and the 12 density functionals should be taken with a strong dose of skepticism.

Beginning by considering the CCSD(T)/CBS results, *trans*-c is predicted to be 0.64 kcal/mol below *cis*-c1 and *cis*-c2 is predicted to be 1.2 kcal/mol above. While the *cis*-1/*trans* difference does not reflect the experimental distribution, the *cis*-2/*cis*-1 difference corresponds to a normalized intensity of 14%, not far removed from the experimental value of 12%. This behavior can be rationalized by considering the closeness of the *cis*-1 and *cis*-2 structures: it is unlikely that further thermodynamic effects play a role in their energy distribution, but the *trans* species has a completely different hydrogen bond network and the difference in entropy will likely play a much bigger role when considering the Gibbs free energies in the next section.

Considering the variability of the *trans* species, we will catalogue the functionals with the closest agreement to the CCSD(T)/CBS *trans*-c/*cis*-c1 energy difference: X3LYP-D3BJ comes the closest with the gap being 0.14 kcal/mol narrower than CCSD(T)/CBS, then CAM-B3LYP-D3BJ (0.17 kcal/mol narrower), then PBE0 (0.2 kcal/mol narrower), B3LYP-D3BJ (0.23 kcal/mol wider than CCSD(T)/CBS), B2PLYP-D3BJ (0.24 kcal/mol narrower). PBE0 is the only functional without dispersion corrections to get within 0.5 kcal/mol of the CCSD(T)/CBS *trans*-c/*cis*-c1 energy difference. Once dispersion corrections are applied (PBE0-D3BJ) the *trans*-c/*cis*-c1 difference widens to 1.2 kcal/mol, nearly twice the value of CCSD(T)/CBS. This observation raises an interesting point: for every case where there is a DFT and DFT-D3BJ pair, the energy of the *trans* conformers relative to the *cis* conformers is decreased when moving from DFT to dispersion-corrected DFT. The favoring of *trans* species by dispersion corrections even holds true across all D3(0)-corrected Minnesota functionals (Figure 4 in green). The effect of dispersion corrections is not always favorable; consider the case of TPSSTPSS wherein *trans* is already considerably favored (2.31 kcal/mol lower than *cis*-1, nearly four times the CCSD(T)/CBS value, shifts to 3.27 kcal/mol at TPSSTPSS-D3BJ), but it is a consistent effect. The unilateral behavior of dispersion corrections indicates that functionals which worsen upon inclusion of dispersion are likely performing well in their uncorrected form due to error compensation and should be treated with skepticism.[45]

Finally, it is important to note the previous work on H^+GPGG. The second pair of bars in the yellow region of Figure 4 presents relative, zero-point corrected, energies of the conformers at the B3LYP/6-311++G(d,p) level of theory, as reported previously.[14] The B3LYP distribution corresponds to normalized counts of 11%, for *cis*-2, and 63% for *trans*, compared to *cis*-1 (note that we report normalized values whereas previous

DOI: 10.1021/acs.jctc.8b00648
J. Chem. Theory Comput. 2018, 14, 5406−5418

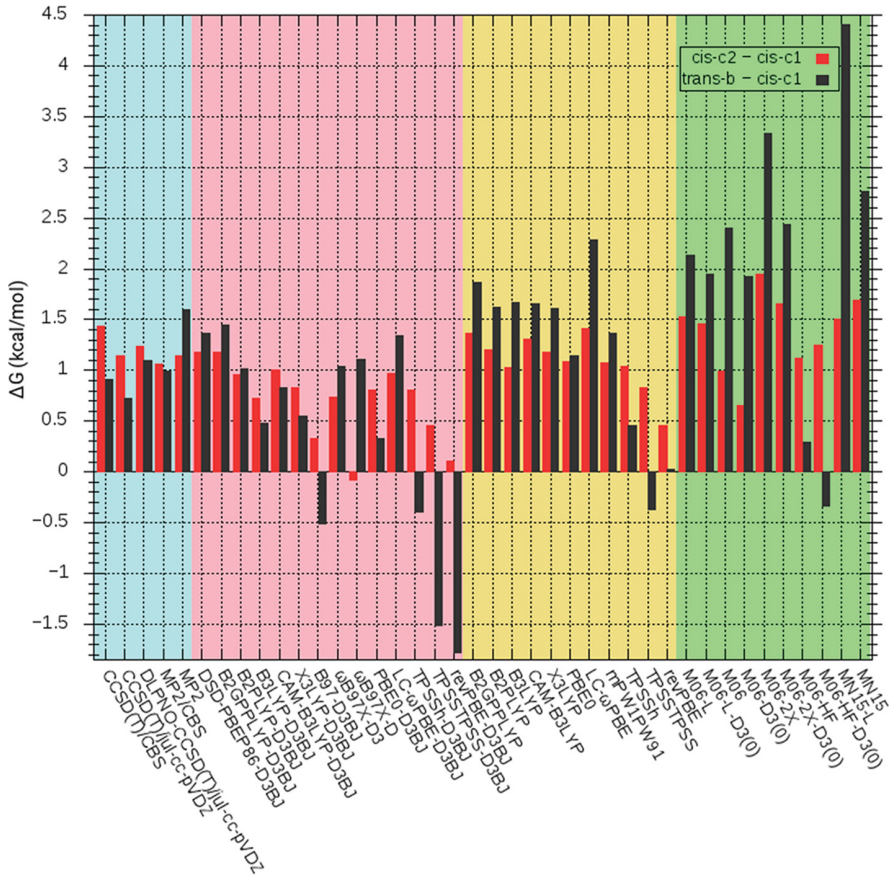


Figure 5. Gibbs free energies of H^+GPGG conformers relative to *cis*-c1. Red bars represent *cis*-c2 relative energies, black bars represent *trans*-b relative energies. CBS values obtained as described in the Methods section. All geometries and vibrational corrections for WFT methods (light blue section) obtained with MP2/6-311++G(d,p), all DFT properties obtained with DFT/6-311++G(d,p). DLPNO with TightPNO cutoffs. Gibbs free energies at 298.15 K, 1 atm.

work reported the unnormalized values, here they are 10% and 39% for *cis*-2 and *trans*, respectively). The conformer distribution follows experiment and, when considering unnormalized values, the values are close enough to experiment to provide apparent verification that the calculated conformers are indeed the observed conformers, however we will show these results to have been obtained via error cancellation once Gibbs free energies are considered in the following section and suggest an alternative method for accurate treatment.

3.5. Gibbs Free Energies, Comparison to Experiment.

Figure 5 reports the Gibbs free energy differences of *cis*-c2 and *trans*-b relative to *cis*-c1. First, it is important to note that the boat form of the *trans* conformer is reported here as it is significantly favored over the chair form. In general, the chair/boat differences for all conformers become more distinct and more unilateral when Gibbs free energies are considered. Of the 40 methods shown in Figure 5, 37 prefer *cis*-c1 over *cis*-b1 (a notable exception being *ω*B97X-D, which favors *cis*-b1 by 0.3 kcal/mol, changing the conformer ordering reported in Figure 5), 33 prefer *cis*-c2 over *cis*-b2, and 30 prefer *trans*-b over *trans*-c. Chair and boat values for every conformer at every level of theory can be found in Table S7 of the Supporting Information.

Aside from the difference in the *trans* boat vs chair preference, it is immediately apparent that the vast majority of methods reported in Figure 5 predict *trans* to be higher in energy than *cis*-1, agreeing with the experimental distribution. CCSD(T)/CBS predicts *trans*-b and *cis*-c2 to be 0.91 and 1.44 kcal/mol higher in energy than *cis*-c1, respectively. The CCSD(T)/CBS distribution corresponds to a *trans* normalized population of 21% and a *cis*-2 population of 9% compared to experimental values of 21% and 12%, respectively. This is excellent agreement confirming the validity of the high level in assessing the experimental distribution as well as the performance of other methods.

Now that comparison to experiment has been established, the menagerie of methods can be whittled down by considering only the methods with the correct conformer ordering with respect to experiment. First, we note the cases where *trans* is lower in energy than *cis*-1: B97-D3BJ, TPSSh-D3BJ, TPSSTPSS-D3BJ, TPSSTPSS, revPBE-D3BJ, revPBE, and M06-HF-D3(0). In these seven cases, *trans*-b was favored too vastly in electronic energies and the unfavorable entropy contribution (vide infra) was unable to place *trans* higher in energy than *cis*-1. As for the ordering of *trans* with respect to *cis*-2, only 10 of the remaining methods place *cis*-2 higher in energy than *trans* to yield the correct ordering with respect to experiment.

Of the 10 methods predicting the correct ordering with respect to experiment, four of them belong to the WFT class discussed previously (CCSD(T)/CBS, CCSD(T)/jul-cc-pVDZ, MP2/CBS, and DLPNO-CCSD(T)/jul-cc-pVDZ with TightPNO cutoffs). As the thermal corrections for all WFT methods come from the same level of theory, the differences between these methods are the same as they were when considering only the electronic energies and will not be discussed further here. Among the tested functionals, four of

DOI: 10.1021/acs.jctc.8b00648
J. Chem. Theory Comput. 2018, 14, 5406−5418

the six with the correct ordering come from the dispersion-corrected class: B3LYP-D3BJ, CAM-B3LYP-D3BJ, X3LYP-D3BJ, and PBE0-D3BJ. B3LYP-D3BJ predicts a normalized *cis*-2 distribution of 30% with respect to *cis*-1 (0.72 kcal/mol lower than CCSD(T)/CBS), and a *trans* distribution of 45% (0.43 kcal/mol lower than CCSD(T)/CBS). While the actual relative energy differences of B3LYP-D3BJ with respect to CCSD(T)/CBS are less than a kcal/mol, the normalized distributions are too high when compared to experiment. Interestingly, the long-range corrected X3LYP-D3BJ performs slightly better than B3LYP-D3BJ with a normalized *cis*-2 distribution of 25% and a *trans* distribution of 40%. PBE0-D3BJ performs slightly worse than X3LYP-D3BJ and B3LYP-D3BJ, with normalized distributions of 25% and 57% for *cis*-2 and *trans*, respectively. The last of the correctly ordered, dispersion-corrected functionals is the range-separated CAM-B3LYP-D3BJ which produces normalized populations of 18% for *cis*-2, compared to 12% in experiment, and 24% for *trans*, compared to 21%. While CAM-B3LYP-D3BJ is the best performing functional out of all tested functionals, these four dispersion-corrected functionals may be considered together in future studies due to their close performance.

Lastly, we should consider the two uncorrected functionals that correctly predict the experimental ordering. TPSSh produces a *cis*-c2 normalized population of 17%, and a *trans*-b population of 46%, comparable with B3LYP-D3BJ. It should be noted, however, that the *trans*-c conformer is 0.08 kcal/mol lower in energy than *trans*-b, yielding a population of 53% if the more favorable *trans* conformer is considered. These results are intriguing, as TPSSh and TPSSTPSS are among the few uncorrected functionals where dispersion corrections worsen the results when compared with CCSD(T)/CBS. Also in this category is M06-HF, the other uncorrected functional with the experimental conformer ordering, yielding populations of 15% and 61% for *cis*-c2 and *trans*-b, respectively. Again, the *trans*-c conformer is favored over *trans*-b in M06-HF; however, here the difference is 0.25 kcal/mol. Given the already close nature of the *trans*-b and *cis*-c1 Gibbs free energies, the population of the *trans* conformer rises to 93% when the lower energy form is considered (an energy difference of 0.04 kcal/mol with respect to *cis*-c1). While the M06-HF *trans* population was already quite high, considering the lower energy *trans* conformer makes M06-HF seem like a poor choice when comparing to experiment, however the performance of TPSSh warrants additional analysis in the future to determine if error cancellation is leading to its favorable, uncorrected, performance.

Both long-range corrections and dispersion corrections are considered boons to density functional theory in general and performed very well in two recent, independent, state of the art benchmarks on single-point energies.[45,76] With this information in mind, it is not surprising that CAM-B3LYP-D3BJ (as well as X3LYP-D3BJ) outperforms B3LYP-D3BJ, and that the dispersion-corrected functionals tend to perform better than the uncorrected functionals. These results also confirm the previous zero-point corrected B3LYP energy result as a false positive and demonstrate that Gibbs free energies need to be considered when calculating the conformational energies of peptides with disparate hydrogen bond networks, especially when the margin for error is as slim as it is here. When comparing these results to other works it should be noted again that all DFT calculations in this work are the result of geometry optimizations and frequency calculations with the specified functional while many test sets are comprised of single-point energies on static structures.[44,45,76,77] While recommendations and trends (such as the good performance of specific classes of functionals or corrections) from these test sets can be useful, the approach presented in this study depends on the coalescence of a number of factors aside from the electronic energy. Fortunately, CAM-B3LYP-D3BJ emerges as a valid alternative to B3LYP when calculating the Gibbs free energy differences of peptide conformers. However, the Gibbs free energy relies on its own set of assumptions, meaning we should analyze the components of the Gibbs free energy and the known variables (scaling of the frequencies and the temperature) to ensure there is no systematic error responsible for these results.

3.6. Deconstructing the Gibbs Free Energy. Figure 6 presents the nonelectronic components of the Gibbs free energy

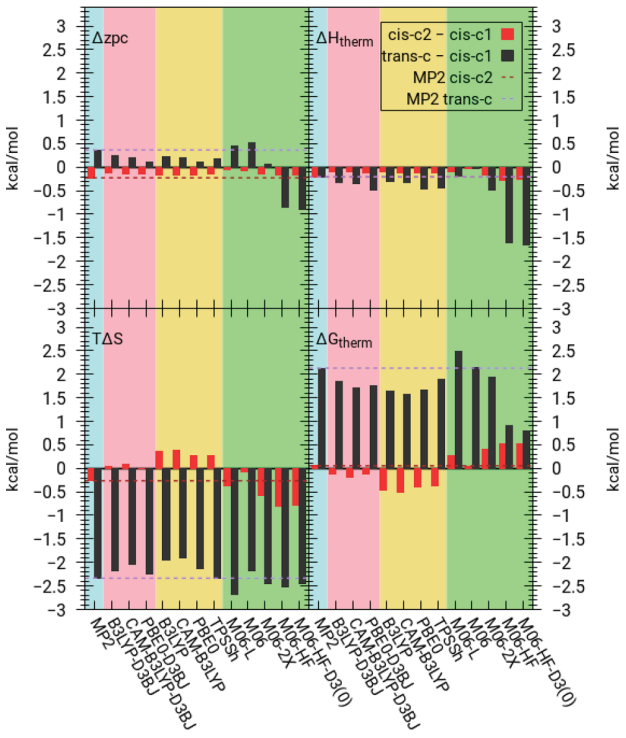


Figure 6. Components of H$^+$GPGG conformer Gibbs free energies relative to *cis*-c1. Zpc corresponds to the zero-point correction, as applied in Figure 4, H_{therm} is the thermal (temperature-dependent) component of the enthalpy, TS is the entropy multiplied by temperature, G_{therm} is the thermal component of the Gibbs free energy. Red bars represent *cis*-c2 relative energies, black bars represent *trans*-c relative energies. Geometries and properties obtained with 6-311+ +G(d,p) basis set, 298.15 K, 1 atm.

differences between the conformers. The reason for investigating the individual components of the free energy is two-fold. First, we need to determine if there is a systematic trend leading to the differences between the zero-point corrected values shown in Figure 4 and the Gibbs free energy values in Figure 5. Second, it is necessary to assess if the differing treatment of the frequencies and geometries plays a role in the reported energy differences between the tested levels of theory. We limit the number of functionals in this step to keep the discussion manageable but include dispersion-corrected and non-dispersion-corrected versions of the majority of the best-performing functionals as well as the span of M06 functionals.

DOI: 10.1021/acs.jctc.8b00648
J. Chem. Theory Comput. 2018, 14, 5406−5418

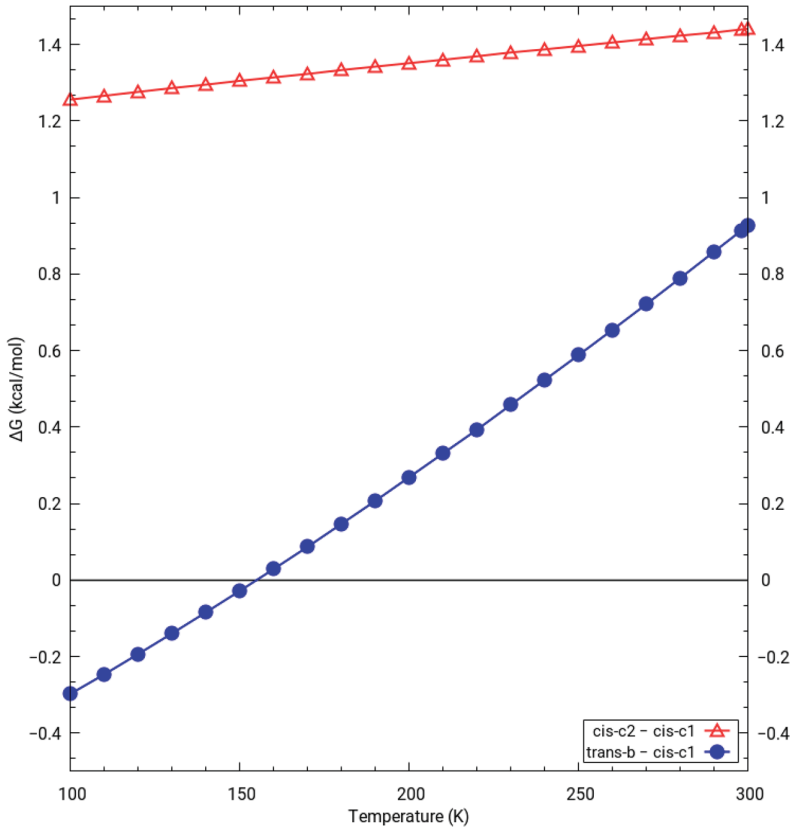


Figure 7. CCSD(T)/CBS//MP2/6-311++G(d,p) Gibbs free energies of H^+GPGG conformers, relative to *cis*-c1, vs temperature.

Starting with the zero-point corrections (Figure 6, upper left), we expect the difference across conformers and levels of theory to be relatively small. The zero-point contribution to a specific conformer energy is proportional to the sum of all frequencies, meaning the zero-point correction difference is a good metric for determining the overall difference in the frequencies produced by a given method or inherent in a given conformer. In general, the difference between conformers is quite small, with the MP2 *cis*-c2 correction being 0.23 kcal/mol smaller than *cis*-c1 and the *trans*-c value falling 0.37 kcal/mol above *cis*-c1. In fact, with the exception of the Minnesota functionals (light green), this behavior is systematic with *cis*-2 lowering in energy with respect to *cis*-1 and *trans* increasing in energy.

The zero-point vibrational energy combined with the translational, rotational, and vibrational contributions to the internal energy yields the thermal enthalpy difference, ΔH_{therm}, shown in the top right panel of Figure 6 (note the PV term in the enthalpy cancels out, making these conformational enthalpy differences identical to internal energy differences, minus the electronic energy). In the majority of cases, the thermal enthalpy differences are less than 0.1 kcal/mol removed from the zero-point energy differences, the exceptions being M06-HF and M06-HF-D03, which lower the ZPVE-only differences by 0.11 kcal/mol each. In every case, the relative *trans* energy is lowered by at least 0.5 kcal/mol, marking the *trans* conformer as enthalpically favored over the *cis* conformers.

The final component leading to the Gibbs free energies is the entropy, presented as $T\Delta S$ in the bottom left panel of Figure 6. Immediately apparent is the unilaterally negative value of the *trans*/*cis*-1 difference. Indeed, *trans* suffers from an unanimously unfavorable entropy contribution compared to *cis*-1, the smallest value being 1.95 kcal/mol (belonging to CAM-B3LYP) and the largest deficit being 2.69 kcal/mol (belonging to M06-L). The consensus among all methods reveals the importance of considering the Gibbs free energy when reporting conformational differences: although the entropy can become a source of error due to the approximations used in attaining it, the fact that 13 methods all predict the *trans* conformer to be entropically disfavored by at least 1.9 kcal/mol cannot be ignored. In fact, the single largest deviation present in Figure 6 comes from the zero-point correction in M06-HF-D3(0) and accounts for the large deviation in the final thermal Gibbs free energy values reported in the bottom right panel of Figure 6. Finally, considering the inclusion of temperature in the Gibbs free energy, we explore the effect of temperature on the conformational distribution.

3.7. Effect of Temperature. Up until this point, Gibbs free energies have been calculated at a temperature of 298.15 K. Room temperature is the general approximation for the temperature of ions in the IMS drift tube; however, there is ongoing debate in the literature on whether the actual temperature of the ions is lower due to the transfer of the ion beam into a low pressure environment.[78,79] Figure 7 reports the CCSD(T)/CBS//MP2/6-311++G(d,p) energies of *cis*-c2 and *trans*-b relative to *cis*-c1, at a range of 100 to 300 K. Throughout all the explored temperatures, the reported chair/boat conformation is the lowest; the crossover of lowest *trans* conformer from *trans*-c to *trans*-b occurs at a temperature lower than 100 K and the gap widens as temperature increases.

As can be inferred from the reported entropies, the *cis*-2 slope with respect to temperature is slight; the energy lowers but not enough to overcome the effects of temperature in the Boltzmann distribution. The normalized *cis*-2 population decreases from 9%

DOI: 10.1021/acs.jctc.8b00648
J. Chem. Theory Comput. 2018, 14, 5406−5418

at 300 K to 0.2% at 100 K, a difference of 0.19 kcal/mol. The *trans* change with temperature is much more drastic, moving from 0.93 kcal/mol less favorable than *cis*-1 (21% normalized population) to 0.30 kcal/mol more favorable. The point where *trans* becomes more favorable than *cis*-1 is 155 K, making us absolutely confident the drift tube temperature is above this value. The experimental *trans* peak corresponds to a normalized population of 21%, meaning the 298.15 K is most likely close to the actual value; however, a deviation of 0.2 kcal/mol would place the *trans* temperature at 270 K (0.72 kcal/mol, 26% normalized population). Certainly, the temperature can be estimated to be above 250 K from these calculations; however, the slim margin for error necessitates further experiments.

4. CONCLUSIONS

A thorough exploration and benchmarking of the low energy conformers of the small, hairpin peptide H^+GPGG has been undertaken with the aid of ion mobility spectrometry. The experimental ion mobility collisional cross section distribution has been reproduced to confirm the results of a wide range of density functionals and wave function theory methods. Experimental IMS peaks were found to correspond to three distinct pairs of conformers, each pair consisting of a chair and boat configuration of the proline. Two of the H^+GPGG conformer pairs possess a *cis* configuration about the Pro–Gly[1] peptide bond and differ by the internal hydrogen-bonding structure, while the other adopts a *trans* configuration. The collisional cross sections were found to be relatively consistent and within the reported 3% error window at every tested level of theory, precluding this quantity as a filter for determining whether a given method is accurate.

While the experimental spectrum reports a higher intensity for the *cis*-1 conformer than the *trans* conformer, 13 WFT levels of theory, including a complete basis set CCSD(T) extapolation, report the *trans* to be favored in electronic energy. When considering the zero-point corrected energies, which are typically reported when considering conformational energies, a similar distribution is encountered. Dispersion-corrected functionals tended to favor the *trans* conformers while uncorrected functionals and the Minnesota family of functionals tended to favor the *cis*-1 conformers. However, upon considering the Gibbs free energies, nearly all functionals favored *cis*-1 over *trans* yet only wave function theory methods and dispersion-corrected density functionals matched the experimental distribution within reasonable error. Of the density functionals tested, the long-range corrected and dispersion-corrected CAM-B3LYP-D3BJ was found to perform most optimally when compared to experiment and CCSD(T)/CBS calculations. These results highlight the importance of not only reporting the Gibbs free energy for systems of disparate hydrogen bond networks but of selecting an appropriate functional for the job. The majority of tested functionals supplied incorrect distributions and even the domain-based local pair-natural orbital (DLPNO) approximation to the CCSD(T) energy was found to produce significant error when tight cutoffs were not imposed.

Additionally, to confirm the validity of considering the Gibbs free energies, we examined the transferability of the thermal components of the Gibbs free energy across various functionals. The considered values (zero-point energies, thermal enthalpies, and entropies) were clustered closely together in the majority of cases; however, the Minnesota functionals were shown to produce some erratic behavior in the vibrational corrections. The entropies were found to highly disfavor the *trans* at every tested level of theory, properly reconciling the electronically favored nature of the *trans* conformers with the experimental distribution and calculated Gibbs free energies. Due to the approximate nature of Gibbs free energy calculations, Gibbs free energies calculated with scaled frequencies were analyzed, and it was found that the relative Gibbs free energy values were largely unaffected by frequency scaling. Finally, the effect of temperature on the calculated distribution was demonstrated, and the population of the *trans* conformer was found to be highly dependent upon temperature. A working range of 270–300 K was suggested to be possible within the IMS drift tube, however further experiments are required to confirm these results.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.8b00648.

Discussion of frequency scaling, DFT prediction of collisional cross sections, effects of DLPNO parameters, chair vs boat distributions, tables containing the relative electronic energies of WFT methods, relative zero-point corrected energies, relative enthalpies, relative Gibbs free energies, and MP2 coordinates, frequencies, and ground state electronic energies. (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: kraghava@indiana.edu.

ORCID

Daniel Beckett: 0000-0003-4833-2269
David E. Clemmer: 0000-0003-4039-1360
Krishnan Raghavachari: 0000-0003-3275-1426

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank Dr. Kevin Gilbert for generous support in the use of PCMODEL as well as stimulating discussion. This work was supported by the Chemical Sciences, Geosciences and Biosciences Division, Office of Basic Energy Sciences, Office of Science, U.S. Department of Energy (DE-FG02-09ER16068) and by the Raymond Siedle Materials Fellowship at Indiana University. Finally, we thank the Big Red II Supercomputing Center at Indiana University for computer time.

■ REFERENCES

(1) Henzler-Wildman, K.; Kern, D. Dynamic personalities of proteins. *Nature* **2007**, *450*, 964–972.

(2) Ruotolo, B. T.; Robinson, C. V. Aspects of native proteins are retained in vacuum. *Curr. Opin. Chem. Biol.* **2006**, *10*, 402–408.

(3) Myung, S.; Badman, E. R.; Lee, Y. J.; Clemmer, D. E. Structural Transitions of Electrosprayed Ubiquitin Ions Stored in an Ion Trap over ~ 10 ms to 30 s. *J. Phys. Chem. A* **2002**, *106*, 9976–9982.

(4) Ruotolo, B. T.; Giles, K.; Campuzano, I.; Sandercock, A. M.; Bateman, R. H.; Robinson, C. V. Evidence for Macromolecular Protein Rings in the Absence of Bulk Water. *Science* **2005**, *310*, 1658.

(5) Pierson, N. A.; Chen, L.; Russell, D. H.; Clemmer, D. E. Cis–Trans Isomerizations of Proline Residues Are Key to Bradykinin Conformations. *J. Am. Chem. Soc.* **2013**, *135*, 3186–3192.

(6) Wyttenbach, T.; Pierson, N. A.; Clemmer, D. E.; Bowers, M. T. Ion Mobility Analysis of Molecular Dynamics. *Annu. Rev. Phys. Chem.* **2014**, *65*, 175–196.

DOI: 10.1021/acs.jctc.8b00648
J. Chem. Theory Comput. 2018, 14, 5406–5418

(7) Hill, H. H.; Siems, W. F.; Louis, R. H. S.; McMinn, D. G. ION MOBILITY SPECTROMETRY. *Anal. Chem.* **1990**, *62*, 1201A−1209A.
(8) Mason, E. A. M. M. *Transport Properties of Ions in Gases*; John Wiley & Sons, Inc.: New York, 1988.
(9) Mack, E. AVERAGE CROSS-SECTIONAL AREAS OF MOLECULES BY GASEOUS DIFFUSION METHODS. *J. Am. Chem. Soc.* **1925**, *47*, 2468−2482.
(10) Shvartsburg, A. A.; Jarrold, M. F. An exact hard-spheres scattering model for the mobilities of polyatomic ions. *Chem. Phys. Lett.* **1996**, *261*, 86−91.
(11) Mesleh, M. F.; Hunter, J. M.; Shvartsburg, A. A.; Schatz, G. C.; Jarrold, M. F. Structural Information from Ion Mobility Measurements: Effects of the Long-Range Potential. *J. Phys. Chem.* **1996**, *100*, 16082−16086.
(12) Saha, A.; Raghavachari, K. Analysis of Different Fragmentation Strategies on a Variety of Large Peptides: Implementation of a Low Level of Theory in Fragment-Based Methods Can Be a Crucial Factor. *J. Chem. Theory Comput.* **2015**, *11*, 2012−2023.
(13) Thapa, B.; Beckett, D.; Jovan Jose, K. V.; Raghavachari, K. Assessment of Fragmentation Strategies for Large Proteins Using the Multilayer Molecules-in-Molecules Approach. *J. Chem. Theory Comput.* **2018**, *14*, 1383−1394.
(14) Masson, A.; Kamrath, M. Z.; Perez, M. A. S.; Glover, M. S.; Rothlisberger, U.; Clemmer, D. E.; Rizzo, T. R. Infrared Spectroscopy of Mobility-Selected H+-Gly-Pro-Gly-Gly (GPGG). *J. Am. Soc. Mass Spectrom.* **2015**, *26*, 1444−1454.
(15) Counterman, A. E.; Clemmer, D. E. Cis−Trans Signatures of Proline-Containing Tryptic Peptides in the Gas Phase. *Anal. Chem.* **2002**, *74*, 1946−1951.
(16) Gilbert, K. *PCMODEL*; Serena Software: Bloomington, IN, 2014.
(17) Halgren, T. A. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490−519.
(18) 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.
(19) Frisch, M. J.; et al. *Gaussian Development Version*, Revision I.06; Gaussian, Inc.: Wallingford, CT, 2016.
(20) Grimme, S. Semiempirical hybrid density functional with perturbative second-order correlation. *J. Chem. Phys.* **2006**, *124*, 034108.
(21) Karton, A.; Tarnopolsky, A.; Lamère, J.-F.; Schatz, G. C.; Martin, J. M. L. Highly Accurate First-Principles Benchmark Data Sets for the Parametrization and Validation of Density Functional and Other Approximate Methods. Derivation of a Robust, Generally Applicable, Double-Hybrid Functional for Thermochemistry and Thermochemical Kinetics. *J. Phys. Chem. A* **2008**, *112*, 12868−12886.
(22) 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.
(23) 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: Condens. Matter Mater. Phys.* **1988**, *37*, 785−789.
(24) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5652.
(25) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Phys. Chem.* **1994**, *98*, 11623−11627.
(26) Adamo, C.; Barone, V. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *J. Chem. Phys.* **1999**, *110*, 6158−6170.
(27) Adamo, C.; Barone, V. Exchange functionals with improved long-range behavior and adiabatic connection methods without adjustable parameters: The mPW and mPW1PW models. *J. Chem. Phys.* **1998**, *108*, 664−675.
(28) Staroverov, V. N.; Scuseria, G. E.; Tao, J.; Perdew, J. P. Comparative assessment of a new nonempirical density functional: Molecules and hydrogen-bonded complexes. *J. Chem. Phys.* **2003**, *119*, 12129−12137.
(29) Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. Climbing the Density Functional Ladder: Nonempirical Meta−Generalized Gradient Approximation Designed for Molecules and Solids. *Phys. Rev. Lett.* **2003**, *91*, 146401.
(30) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.
(31) Zhang, Y.; Yang, W. Comment on ̀Generalized Gradient Approximation Made Simple". *Phys. Rev. Lett.* **1998**, *80*, 890−890.
(32) Zhao, Y.; Truhlar, D. G. A new local density functional for main-group thermochemistry, transition metal bonding, thermochemical kinetics, and noncovalent interactions. *J. Chem. Phys.* **2006**, *125*, 194101.
(33) 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.
(34) Zhao, Y.; Truhlar, D. G. Density Functional for Spectroscopy: No Long-Range Self-Interaction Error, Good Performance for Rydberg and Charge-Transfer States, and Better Performance on Average than B3LYP for Ground States. *J. Phys. Chem. A* **2006**, *110*, 13126−13130.
(35) Yu, H. S.; He, X.; Truhlar, D. G. MN15-L: A New Local Exchange-Correlation Functional for Kohn−Sham Density Functional Theory with Broad Accuracy for Atoms, Molecules, and Solids. *J. Chem. Theory Comput.* **2016**, *12*, 1280−1293.
(36) Yu, H. S.; He, X.; Li, S. L.; Truhlar, D. G. MN15: A Kohn-Sham global-hybrid exchange-correlation density functional with broad accuracy for multi-reference and single-reference systems and non-covalent interactions. *Chem. Sci.* **2016**, *7*, 5032−5051.
(37) Yanai, T.; Tew, D. P.; Handy, N. C. A new hybrid exchange−correlation functional using the Coulomb-attenuating method (CAM-B3LYP). *Chem. Phys. Lett.* **2004**, *393*, 51−57.
(38) Vydrov, O. A.; Scuseria, G. E. Assessment of a long-range corrected hybrid functional. *J. Chem. Phys.* **2006**, *125*, 234109.
(39) Chai, J.-D.; Head-Gordon, M. Long-range corrected hybrid density functionals with damped atom-atom dispersion corrections. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615−6620.
(40) Lin, Y.-S.; Li, G.-D.; Mao, S.-P.; Chai, J.-D. Long-Range Corrected Hybrid Density Functionals with Improved Dispersion Corrections. *J. Chem. Theory Comput.* **2013**, *9*, 263−272.
(41) Xu, X.; Goddard, W. A. The X3LYP extended density functional for accurate descriptions of nonbond interactions, spin states, and thermochemical properties. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 2673−2677.
(42) 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.
(43) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011**, *32*, 1456−1465.
(44) Goerigk, L.; Grimme, S. A thorough benchmark of density functional methods for general main group thermochemistry, kinetics, and noncovalent interactions. *Phys. Chem. Chem. Phys.* **2011**, *13*, 6670−6688.
(45) Goerigk, L.; Hansen, A.; Bauer, C.; Ehrlich, S.; Najibi, A.; Grimme, S. A look at the density functional theory zoo with the advanced GMTKN55 database for general main group thermochemistry, kinetics and noncovalent interactions. *Phys. Chem. Chem. Phys.* **2017**, *19*, 32184−32215.
(46) Grimme, S. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem.* **2006**, *27*, 1787−1799.

DOI: 10.1021/acs.jctc.8b00648
J. Chem. Theory Comput. 2018, 14, 5406−5418

(47) Kozuch, S.; Martin, J. M. L. DSD-PBEP86: in search of the best double-hybrid DFT with spin-component scaled MP2 and dispersion corrections. *Phys. Chem. Chem. Phys.* **2011**, *13*, 20104−20107.
(48) Kozuch, S.; Martin, J. M. L. Spin-component-scaled double hybrids: An extensive search for the best fifth-rung functionals blending DFT and perturbation theory. *J. Comput. Chem.* **2013**, *34*, 2327−2344.
(49) Frisch, M. J.; et al. *Gaussian 16*, revision A.01; Gaussian, Inc.: Wallingford, CT, 2016.
(50) Neese, F. The ORCA program system. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012**, *2*, 73−78.
(51) Frisch, M. J.; Pople, J. A.; Binkley, J. S. Self-consistent molecular orbital methods 25. Supplementary functions for Gaussian basis sets. *J. Chem. Phys.* **1984**, *80*, 3265−3269.
(52) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions. *J. Chem. Phys.* **1980**, *72*, 650−654.
(53) Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. V. R. Efficient diffuse function-augmented basis sets for anion calculations. III. The 3-21+G basis set for first-row elements, Li−F. *J. Comput. Chem.* **1983**, *4*, 294−301.
(54) Feyereisen, M.; Fitzgerald, G.; Komornicki, A. Use of approximate integrals in ab initio theory. An application in MP2 energy calculations. *Chem. Phys. Lett.* **1993**, *208*, 359−363.
(55) Ten-no, S.; Manby, F. R. Density fitting for the decomposition of three-electron integrals in explicitly correlated electronic structure theory. *J. Chem. Phys.* **2003**, *119*, 5358−5363.
(56) Papajak, E.; Zheng, J.; Xu, X.; Leverentz, H. R.; Truhlar, D. G. Perspectives on Basis Sets Beautiful: Seasonal Plantings of Diffuse Basis Functions. *J. Chem. Theory Comput.* **2011**, *7*, 3027−3034.
(57) Kendall, R. A.; Dunning, T. H.; Harrison, R. J. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *J. Chem. Phys.* **1992**, *96*, 6796−6806.
(58) Dunning, 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.
(59) Jurečka, P.; Hobza, P. On the convergence of the (ΔECCSD(T)−ΔEMP2) term for complexes with multiple H-bonds. *Chem. Phys. Lett.* **2002**, *365*, 89−94.
(60) Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. Basis-set convergence in correlated calculations on Ne, N2, and H2O. *Chem. Phys. Lett.* **1998**, *286*, 243−252.
(61) Riplinger, C.; Neese, F. An efficient and near linear scaling pair natural orbital based local coupled cluster method. *J. Chem. Phys.* **2013**, *138*, 034106.
(62) 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.
(63) 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.
(64) Merrick, J. P.; Moran, D.; Radom, L. An Evaluation of Harmonic Vibrational Frequency Scale Factors. *J. Phys. Chem. A* **2007**, *111*, 11683−11700.
(65) Merenbloom, S. I.; Koeniger, S. L.; Valentine, S. J.; Plasencia, M. D.; Clemmer, D. E. IMS−IMS and IMS−IMS−IMS/MS for Separating Peptide and Protein Fragment Ions. *Anal. Chem.* **2006**, *78*, 2802−2809.
(66) Tang, K.; Shvartsburg, A. A.; Lee, H.-N.; Prior, D. C.; Buschbach, M. A.; Li, F.; Tolmachev, A. V.; Anderson, G. A.; Smith, R. D. High-Sensitivity Ion Mobility Spectrometry/Mass Spectrometry Using Electrodynamic Ion Funnel Interfaces. *Anal. Chem.* **2005**, *77*, 3330−3339.
(67) Pierson, N. A.; Valentine, S. J.; Clemmer, D. E. Evidence for a quasi-equilibrium distribution of states for bradykinin [M + 3H]3+ ions in the gas phase. *J. Phys. Chem. B* **2010**, *114*, 7777−7783.
(68) Jeffrey, G. A. *An Introduction to Hydrogen Bonding*; Oxford University Press: Oxford, U.K., 1997.
(69) Goerigk, L.; Reimers, J. R. Efficient Methods for the Quantum Chemical Treatment of Protein Structures: The Effects of London-Dispersion and Basis-Set Incompleteness on Peptide and Water-Cluster Geometries. *J. Chem. Theory Comput.* **2013**, *9*, 3240−3251.
(70) Goerigk, L.; Collyer, C. A.; Reimers, J. R. Recommending Hartree−Fock Theory with London-Dispersion and Basis-Set-Superposition Corrections for the Optimization or Quantum Refinement of Protein Structures. *J. Phys. Chem. B* **2014**, *118*, 14612−14626.
(71) DeTar, D. F.; Luthra, N. P. Conformations of proline. *J. Am. Chem. Soc.* **1977**, *99*, 1232−1244.
(72) Chakrabarti, P.; Pal, D. The interrelationships of side-chain and main-chain conformations in proteins. *Prog. Biophys. Mol. Biol.* **2001**, *76*, 1−102.
(73) Ho, B. K. The flexibility in the proline ring couples to the protein backbone. *Protein Sci.* **2005**, *14*, 1011−1018.
(74) Siu, C.-K.; Guo, Y.; Saminathan, I. S.; Hopkinson, A. C.; Siu, K. W. M. Optimization of Parameters Used in Algorithms of Ion-Mobility Calculation for Conformational Analyses. *J. Phys. Chem. B* **2010**, *114*, 1204−1212.
(75) Goerigk, L.; Karton, A.; Martin, J. M. L.; Radom, L. Accurate quantum chemical energies for tetrapeptide conformations: why MP2 data with an insufficient basis set should be handled with caution. *Phys. Chem. Chem. Phys.* **2013**, *15*, 7028−7031.
(76) Mardirossian, N.; Head-Gordon, M. Thirty years of density functional theory in computational chemistry: an overview and extensive assessment of 200 density functionals. *Mol. Phys.* **2017**, *115*, 2315−2372.
(77) Goerigk, L.; Grimme, S. A General Database for Main Group Thermochemistry, Kinetics, and Noncovalent Interactions − Assessment of Common and Reparameterized (meta-)GGA Density Functionals. *J. Chem. Theory Comput.* **2010**, *6*, 107−126.
(78) Lee, S.-W.; Freivogel, P.; Schindler, T.; Beauchamp, J. L. Freeze-Dried Biomolecules: FT-ICR Studies of the Specific Solvation of Functional Groups and Clathrate Formation Observed by the Slow Evaporation of Water from Hydrated Peptides and Model Compounds in the Gas Phase. *J. Am. Chem. Soc.* **1998**, *120*, 11758−11765.
(79) Rodriguez-Cruz, S. E.; Klassen, J. S.; Williams, E. R. Hydration of gas-phase gramicidin S (M + 2H)2+ ions formed by electrospray: The transition from solution to gas-phase structure. *J. Am. Soc. Mass Spectrom.* **1997**, *8*, 565−568.

DOI: 10.1021/acs.jctc.8b00648
J. Chem. Theory Comput. 2018, 14, 5406−5418