

Tyrosine-glycine revisited: Resolving the discrepancy between theory and experiment

Leo F. Holroyd and Tanja van Mourik*

EaStCHEM School of Chemistry, University of St Andrews. North Haugh, St Andrews KY16 9ST (UK)

Corresponding author:

Tanja van Mourik

Email: tanja.vanmourik@st-andrews.ac.uk

Phone: +44 (0)1334 463822

Fax: +44 (0)1334 463808

Abstract

Energies of 20 conformers of the Tyr-Gly dipeptide were computed using DSD-PBEP86-D3BJ/aug-cc-VTZ, with geometries from M06-2X/6-31+G* and B97-D/6-31+G*. At 0 K, these energies support the earlier finding from MP2/6-31+G**/B3LYP/6-31+G*, that the most stable conformer is folded and H-bonded. However, when free-energy corrections at 400 K are added, non-H-bonded conformers are the most stable. This supports an earlier spectroscopic study in which H-bonded conformers were absent. Of the four most stable conformers at 400 K, two were not matched with spectra in the experimental study, but we argue that all four can in fact be plausibly assigned to the experimental spectra.

1. Introduction

The neutral dipeptide tyrosine-glycine (Tyr-Gly) displays a high degree of conformational flexibility: it can be folded or extended, and an intramolecular hydrogen bond between OH(Gly) and C=O(Tyr) can be present or absent. In 2006, our complete search of the conformational space of this molecule – generating all possible conformers, and sorting them by their 0 K gas-phase potential energies at successively higher levels of theory – found the global minimum (“conformer 1”) to be a folded, H-bonded structure at the ZPE-corrected MP2/6-31+G**/B3LYP/6-31+G* level of theory [1]. The 20 most stable conformers from that study are shown in Figure 1. The top six structures are folded, and the rest are extended. At 0 K, the preference for conformer 1 is easy to rationalise: folding allows the Tyr aromatic side chain to interact dispersively with the backbone, while the H-bond further lowers the potential energy. Understanding the behaviour of gas-phase peptides has important implications for biomolecular structure – although proteins are solvated in the cell, the residues of their hydrophobic cores are expected to experience an essentially gas-phase environment [2].

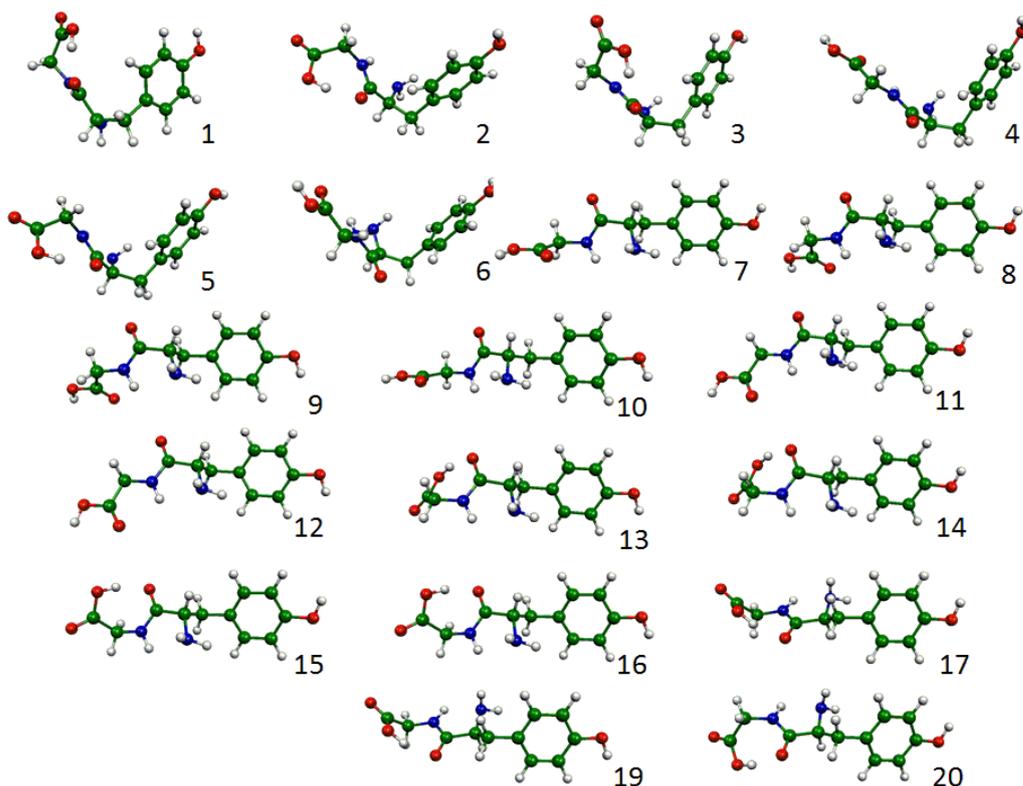


Figure 1. The 20 most stable conformers of Tyr-Gly at the ZPE-corrected MP2/6-31+G**/B3LYP/6-31+G* level of theory, from Ref. [1]. Conformers 1–6 are folded; conformers 1, 2, 3, 5 and 13–20 are H-bonded.

In 2011, the spectroscopic group of de Vries performed resonant two-photon ionization (R2PI) experiments on gas-phase Tyr-Gly, obtaining four UV–UV and IR–UV double resonance spectra [3]. These spectra were compared with the calculated vibrational frequencies of a set of newly optimised Tyr-Gly energy minima, in order to identify the conformers detected in the experiment. Neither the hydroxyl nor the carboxylic O-H stretching frequencies in the IR–UV spectra were red-shifted from the typical free hydroxyl value, indicating that they were not involved in H-bonding, and thus ruling out conformer 1. Guided by B3LYP-calculated frequencies, the authors in fact assigned their four recorded spectra to four computed gas-phase structures which were extended and not H-bonded. These conformers had also been obtained in our study (wherein they were numbered 7, 10, 11 and 12, cf. Figure 1), but in our study they were not the most stable in terms of potential energy: they lay ~ 6 kJ mol⁻¹ above conformer 1 at 0 K [1]. Thus theory and experiment appeared to disagree.

The experimental absence of peptide conformers that are predicted by theory is a well-attested phenomenon [4-7]. In this Letter, we apply newer, more robust DFT methods to address this problem. We also perform calculations to explore the possibility that entropic effects play a role – although R2PI spectra are measured at supercool temperatures, the sample might approximately preserve the Boltzmann distribution of its high formation temperature if the cooling is a non-equilibrium process [6]. The importance of thermodynamic effects has been observed before. Piquemal, Mons *et al.* found that the coexistence of a folded and extended conformer of the small peptide *N*-acetyl-(Ala)₂-O-benzyl, as observed experimentally, can be rationalised through their Gibbs free energies.

The authors argue that thermodynamic corrections need to be taken into account to computationally locate the most stable conformer [8]. In light of our new results, we reconsider the original spectral assignments made by de Vries *et al.* We also mention recent experimental findings which highlight the possible role of non-radiative deactivation of excited states of folded peptide conformers.

2. Computational methods

Gibbs free energy corrections, based on the vibrational entropy S_{vib} calculated within the harmonic approximation, were initially calculated at the B3LYP/6-31+G* level of theory for the 20 lowest-energy conformers of Tyr-Gly optimised with that method in Ref. [1]. These corrections were calculated at two temperatures, 298.15 K and 400 K, using Gaussian 09 [9].

However, the B3LYP functional might well be unreliable for peptides with aromatic side chains, since it lacks dispersion and presumably underestimates folding [2,5,10,11]. Furthermore, MP2 (used for the final single-point energy ordering) is susceptible to an intramolecular form of the basis set superposition error (BSSE), which unfairly *favours* folded structures [10,12,13]. Though they are opposite in sign, these two errors cannot be assumed to cancel simply by combining B3LYP geometries with MP2 energies [11]. Therefore, to obtain more reliable geometries – accounting for dispersion while avoiding BSSE – we re-optimised the 20 most stable conformers from Ref. [1] with B97-D/6-31+G* and M06-2X/6-31+G*. These functionals were previously identified by Piquemal, Mons *et al.* as suitable for the treatment of peptide folding, in which dispersion is crucial [8]. They exemplify two different empirical approaches to recovering the missing dispersion in DFT – the former [14] by including an atom-pairwise dispersion correction of the form $C_6 \cdot R^{-6}$, the latter [15] through parameters derived from training sets containing dispersion-bound complexes. DFT in general is largely free of BSSE [16]. Gibbs free energy corrections for the newly-optimised structures were also calculated at 298.15 K and 400 K using the corresponding model chemistries. These calculations were also performed with Gaussian 09 [9].

Single-point energies of these structures were then calculated at the DSD-PBEP86-D3BJ/aug-cc-pVTZ level of theory, implemented in ORCA 3.0 [17]. This functional [18], which belongs to the DSD class of methods [19] (containing a fraction of spin-component-scaled MP2 correlation energy as well as Hartree-Fock exchange, plus a D3BJ-type Grimme dispersion term, which we calculated with the program DFT-D3 [20]) has performed promisingly well in general benchmark tests [18,21], and specifically for dipeptides [22]. The choice of a relatively large basis set should help to minimise the BSSE, an error to which DSD functionals are probably more susceptible than standard DFT methods due to the inclusion of MP2 correlation.

3. Results and discussion

Adding B3LYP/6-31+G* free energy corrections to the MP2/6-31+G* single-points induces a preference for conformers 7 and 10 (Table 1). However, the effect is modest: these conformers now lie approximately 0.8 kJ mol^{-1} (at 298.15 K) or 1.8 kJ mol^{-1} (400 K) below the lowest-lying folded conformer (number 4). The relative populations of the 20 conformers at those temperatures, calculated according to a Maxwell–Boltzmann distribution (Table 1),

suggest that at least one of the folded conformers (1–6) should be found, as well as 7 and 10. Also, a clear effect of the entropic correction is to heavily disfavour extended conformers with a hydrogen bond (numbers 13–20).

Table 1

MP2/6-31+G**/B3LYP/6-31+G* Gibbs free energies of 20 conformers of Tyr-Gly, relative to conformer 1, at 298.15 and 400 K, in kJ mol⁻¹. In brackets are the Maxwell–Boltzmann relative populations.

| Conf. | ΔG , 298.15 K (pop) | ΔG , 400 K (pop) | Conf. | ΔG , 298.15 K (pop) | ΔG , 400 K (pop) |
|-------|--------------------------------|-----------------------------|-------|--------------------------------|-----------------------------|
| 1 | 0.00 (584) | 0.00 (271) | 11 | 0.93 (401) | -1.68 (450) |
| 2 | 4.13 (110) | 4.38 (73) | 12 | 1.07 (380) | -1.52 (428) |
| 3 | 3.17 (163) | 2.97 (111) | 13 | 6.38 (44) | 6.04 (44) |
| 4 | -0.49 (713) | -2.57 (587) | 14 | 6.65 (40) | 6.38 (40) |
| 5 | 4.78 (85) | 4.97 (61) | 15 | 6.71 (39) | 6.41 (40) |
| 6 | 0.31 (516) | -1.74 (457) | 16 | 6.77 (38) | 6.43 (39) |
| 7 | -1.33 (1000) | -4.34 (1000) | 17 | 8.48 (19) | 7.60 (28) |
| 8 | -0.01 (586) | -2.53 (581) | 18 | 9.12 (15) | 8.44 (21) |
| 9 | 0.22 (534) | -2.25 (534) | 19 | 8.54 (29) | 7.68 (27) |
| 10 | -1.18 (939) | -4.22 (964) | 20 | 9.81 (11) | 9.25 (17) |

When the 20 conformers were re-optimised with B97-D/6-31+G* and M06-2X/6-31+G*, they all became more folded, as quantified by a decrease in the distance between the C_{carb}(Gly) and C1(Tyr) atoms. For example, this distance decreases from 6.3 to 5.3 Å using M06-2X.

Next we consider the double hybrid functional. In terms of 0 K single-point energies, DSD-PBEP86-D3BJ/aug-cc-pVTZ favours conformer 1 over the extended conformers by 7 kJ mol⁻¹ or more, whether using the B97-D/6-31+G* or the M06-2X/6-31+G* geometries (Table 2).

Table 2

DSD-PBEP86-D3BJ/aug-cc-pVTZ//B97-D/6-31+G* and DSD-PBEP86-D3BJ/aug-cc-pVTZ//M06-2X/6-31+G* potential energies of 20 conformers of Tyr-Gly, relative to conformer 1, in kJ mol^{-1} .

| Conf. | B97-D | M06-2X | Conf. | B97-D | M06-2X |
|-------|-------|--------|-------|-------|--------|
| 1 | 0.00 | 0.00 | 11 | 11.08 | 10.47 |
| 2 | 3.96 | 3.83 | 12 | 11.11 | 10.51 |
| 3 | 3.37 | 3.22 | 13 | 7.25 | 7.00 |
| 4 | 8.16 | 7.78 | 14 | 7.34 | 6.96 |
| 5 | 4.62 | 4.46 | 15 | 7.29 | 6.91 |
| 6 | 8.65 | 8.23 | 16 | 7.42 | 7.09 |
| 7 | 11.02 | 10.98 | 17 | 10.96 | 10.85 |
| 8 | 10.88 | 10.49 | 18 | 11.28 | 11.14 |
| 9 | 10.86 | 10.43 | 19 | 10.99 | 10.88 |
| 10 | 11.18 | 11.13 | 20 | 11.57 | 11.38 |

However, when the free energy corrections from B97-D or M06-2X are included, the most stable extended conformers (11 and 12) become essentially isoenergetic with the most stable folded conformers (1, 4 and 6) at 298.15 and 400 K (Tables 3 and 4). From the resulting Boltzmann distributions (Tables 3 and 4), it is clear that, at either temperature's equilibrium, both folded and extended conformers should be present. At the lower temperature, the folded conformers should include one with an H-bond (conformer 1) as well as two without (conformers 4 and 6), but at the higher temperature, probably no H-bonded structures should be found. Extended, non-H-bonded conformers (11 and 12) are expected at both temperatures. The discrepancy between theory and experiment therefore remains: conformers 4, 6 and (possibly) 1 are expected, but they were not observed.

Table 3

DSD-PBEP86-D3BJ/aug-cc-pVTZ//B97-D/6-31+G* Gibbs free energies of 20 conformers of Tyr-Gly, relative to conformer 1, at 298.15 and 400 K, in kJ mol^{-1} . In brackets are the relative Maxwell-Boltzmann populations.

| Conf. | ΔG , 298.15 K (pop) | ΔG , 400 K (pop) | Conf. | ΔG , 298.15 K (pop) | ΔG , 400 K (pop) |
|-------|--------------------------------|-----------------------------|-------|--------------------------------|-----------------------------|
| 1 | 0.00 (906) | 0.00 (380) | 11 | 0.16 (848) | -3.22 (1000) |
| 2 | 4.12 (172) | 4.11 (110) | 12 | 0.17 (844) | -3.20 (995) |
| 3 | 2.70 (305) | 2.49 (180) | 13 | 5.13 (114) | 4.40 (101) |
| 4 | -0.25 (1000) | -2.81 (885) | 14 | 4.92 (124) | 4.14 (109) |
| 5 | 4.56 (144) | 4.50 (98) | 15 | 5.28 (107) | 4.56 (96) |
| 6 | -0.03 (918) | -2.67 (847) | 16 | 5.32 (106) | 4.57 (96) |
| 7 | 2.70 (305) | -0.03 (383) | 17 | 6.98 (54) | 5.72 (68) |
| 8 | 2.47 (334) | -0.26 (410) | 18 | 7.48 (44) | 6.26 (58) |
| 9 | 2.34 (352) | -0.41 (430) | 19 | 6.83 (58) | 5.52 (72) |
| 10 | 2.64 (312) | -0.16 (398) | 20 | 7.57 (43) | 6.30 (57) |

Table 4

DSD-PBEP86-D3BJ/aug-cc-pVTZ//M06-2X/6-31+G* Gibbs free energies of 20 conformers of Tyr-Gly, relative to conformer 1, at 298.15 and 400 K, in kJ mol⁻¹. In brackets are the relative Maxwell–Boltzmann populations.

| Conf. | ΔG , 298.15 K (pop) | ΔG , 400 K (pop) | Conf. | ΔG , 298.15 K (pop) | ΔG , 400 K (pop) |
|-------|--------------------------------|-----------------------------|-------|--------------------------------|-----------------------------|
| 1 | 0.00 (1000) | 0.00 (476) | 11 | 0.47 (829) | -2.47 (1000) |
| 2 | 3.17 (278) | 2.88 (200) | 12 | 0.56 (798) | -2.38 (971) |
| 3 | 1.56 (533) | 1.03 (348) | 13 | 3.71 (224) | 2.57 (220) |
| 4 | 0.27 (895) | -1.86 (833) | 14 | 4.34 (174) | 3.39 (172) |
| 5 | 3.42 (251) | 3.03 (191) | 15 | 4.45 (166) | 3.56 (163) |
| 6 | 1.13 (633) | -0.89 (621) | 16 | 4.62 (155) | 3.73 (155) |
| 7 | 1.89 (467) | -1.00 (643) | 17 | 6.78 (65) | 5.47 (92) |
| 8 | 2.92 (308) | 0.59 (398) | 18 | 7.08 (58) | 5.75 (84) |
| 9 | 3.13 (283) | 0.89 (364) | 19 | 6.94 (61) | 5.66 (87) |
| 10 | 2.63 (347) | -0.11 (492) | 20 | 7.46 (49) | 6.18 (74) |

These results suggest that the absence of conformer 1 from the experimental study in Ref. [3] might have been due to a high formation temperature of the compound. If the Tyr-Gly sample was formed (or subsequently laser-desorbed) at a temperature closer to 400 K than to ambient conditions, hydrogen bonding would have been entropically disfavoured, and conformer 1 would not have been among the four most populated states, assuming a Maxwell–Boltzmann distribution.

Alternatively, the apparent absence of conformer 1 may have been due to a short excited-state lifetime. Highly efficient non-radiative relaxation pathways for excited states of hydrogen-bonded peptide conformers were recently proposed by Mališ *et al.* [23] A folded conformer of *N*-acetylphenylalaninylamide (NAPA), containing two intramolecular H-bonds (one of the NH•••O type, the other an NH– π bond between the N_{Phe}H bond and the phenyl ring), was revealed by R2PI experiments to have a short excited-state lifetime (~1.5 ns). This explained its absence from earlier laser-induced fluorescence studies. Time-dependent DFT-based molecular dynamics simulations (TD-DFT MD) indicated three possible ultrafast deactivation mechanisms, based on either N_{Phe}-H bond stretching or backbone amide distortions. Domcke and Sobolewski subsequently argued that the mechanism involving H-atom transfer along the NH•••O hydrogen bond was the most plausible, as it had the lowest barrier and did not involve the motion of heavy atoms [7]; based on structural similarity, this mechanism might also conceivably operate in the OH•••O hydrogen bond of Tyr-Gly conformer 1.

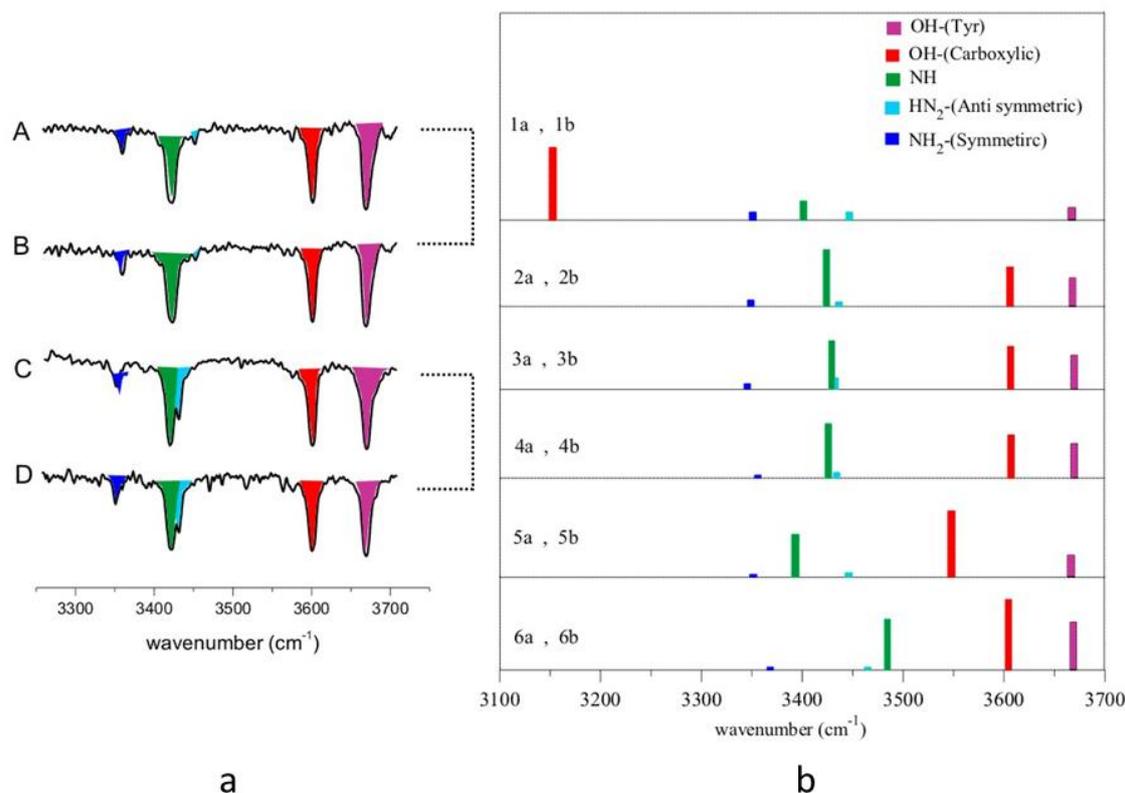


Figure 2. The vibrational spectra of Tyr-Gly. a. Recorded IR–UV spectra. b. Calculated spectra from B3LYP/6-31G**, scaled by 0.9602. Reprinted (adapted) with permission from A. Abo-Riziq, L. Grace, B. Crews, M.P. Callahan, T. van Mourik and M.S. de Vries, *J. Phys. Chem. A* 115, 6077-6087 (2011). Copyright 2011 American Chemical Society.

Meanwhile, the apparent absence of conformers 4 and 6 from experiment [3] can be resolved if the spectra are interpreted differently. The calculated spectra of those conformers (labelled 2a and 2b in Ref. [3]) were extremely similar to the calculated spectra of the conformers labelled 3a and 3b in that study, but the latter were judged to more closely match the experimental spectra labelled C and D (Figure 2). If C and D are instead assigned to 2a and 2b (i.e. 4 and 6), then the total set of structures observed in that study becomes 4, 6, 11 and 12. At 400 K, these are the four most stable conformers at the DSD-PBEP86-D3BJ/aug-cc-pVTZ//B97-D/6-31+G* level of theory (Table 3). With the M06-2X/6-31+G* geometries (Table 4), conformer 6 is isoenergetic with conformer 7, which was not considered in Ref. [3]. However, that study used B3LYP/6-31G** – when we optimised conformer 7 with that model chemistry, it became conformer 11. It therefore seems highly likely that conformers 4, 6, 11 and 12, which are the high-level theoretically predicted structures at 400 K, were observed by de Vries *et al.*

4. Conclusion

When sources of computational error (BSSE and missing dispersion) are minimised, the predicted structural diversity of gas-phase tyrosine-glycine at 0 K is slightly broader than

what is observed experimentally. However, thermal effects should not be neglected. If the population distribution of Tyr-Gly conformers is calculated at 400 K, with free-energy corrections added to the results obtained by high-level DFT methods that account for electron correlation, the result is consistent with experiment, when a different (but equally plausible) choice of spectral assignments is made from the earlier R2PI study. At 400 K, theory predicts a range of non-H-bonded conformers, namely a mixture of those that are folded (4 and 6) and those that are extended (11 and 12) – and that is arguably what was observed experimentally. The presence of the folded conformers 4 and 6, though not originally assigned by de Vries *et al.*, is perfectly consistent with their spectra.

Meanwhile, conformer 1, predicted to be the gas-phase global minimum by Toroz and van Mourik, was absent from experiment because of the entropic bias against H-bonding at high temperatures. This conformer might also have been impossible to observe anyway because of fast non-radiative decay of its excited state; however, since the experiments by de Vries *et al.* used 200 ps (0.2 ns) pulses, and the fast decay revealed by Mališ *et al.* for NAPA resulted in a lifetime of ~1.5 ns, this explanation seems less plausible. We therefore propose that thermal effects are the more likely explanation for the absence of conformer 1, and that conformers 4, 6, 11 and 12 were observed in the experiment, consistent with high-level theory. However, future experiments may be helpful here. We also hope that this study has usefully highlighted several potential difficulties in calculating gas-phase peptide properties and using them to interpret experimental results.

Acknowledgements

The authors gratefully acknowledge EaStCHEM for computer time on the EaStCHEM Research Computing Facility. LFH is grateful to the Engineering and Physical Sciences Research Council for studentship support through the Doctoral Training Account scheme.

References

- [1] T. van Mourik, D. Toroz, *Mol. Phys.* 104 (2006) 559.
- [2] H. Valdés, K. Pluhackova, P. Hobza, *J. Chem. Theory Comput.* 5 (2009) 2248.
- [3] A. Abo-Riziq, L. Grace, B. Crews, M.P. Callahan, T. van Mourik, M.S. de Vries, *J. Phys. Chem. A* 115 (2011) 6077.
- [4] P.D. Godfrey, R.D. Brown, F.M. Rodgers, *J. Mol. Struct.* 376 (1996) 65.
- [5] D. Řeha, H. Valdés, J. Vondrášek, P. Hobza, A. Abu-Riziq, B. Crews, M.S. De Vries, *Chem. Eur. J.* 11 (2005) 6803.
- [6] H. Valdés, V. Spiwok, J. Rezac, D. Reha, A. Abo-Riziq, M.S. De Vries, P. Hobza, *Chem. Eur. J.* 14 (2008) 4886.
- [7] W. Domcke, A.L. Sobolewski, *Nat. Chem.* 5 (2013) 257.
- [8] E. Gloaguen, B. de Courcy, J.P. Piquemal, J. Pilmé, O. Parisel, R. Pollet, H.S. Biswal, F. Piuze, B. Tardivel, M. Broquier, M. Mons, *J. Am. Chem. Soc.* 132 (2010) 11860.
- [9] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G.A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H.P. Hratchian, A.F. Izmaylov, J. Bloino, G. Zheng, J.L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J.A. Montgomery Jr., J.E. Peralta, F. Ogliaro, M. Bearpark, J.J. Heyd, E. Brothers, K.N. Kudin, V.N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J.M. Millam, M. Klene, J.E. Knox, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W.

- Ochterski, R.L. Martin, K. Morokuma, V.G. Zakrzewski, G.A. Voth, P. Salvador, J.J. Dannenberg, S. Dapprich, A.D. Daniels, Ö. Farkas, J.B. Foresman, J.V. Ortiz, J. Cioslowski, D.J. Fox, Gaussian 09, Revision A.02. Gaussian, Inc., Wallingford CT, 2009.
- [10] L.F. Holroyd, T. van Mourik, Chem. Phys. Lett. 442 (2007) 42.
- [11] H. Kruse, L. Goerigk, S. Grimme, J. Org. Chem. 77 (2012) 10824.
- [12] T. van Mourik, P.G. Karamertzanis, S.L. Price, J. Phys. Chem. A 110 (2006) 8.
- [13] T. van Mourik, A.E. Shields, J. Phys. Chem. A 111 (2007) 13272.
- [14] S. Grimme, J. Comput. Chem. 27 (2006) 1787.
- [15] Y. Zhao, D.G. Truhlar, Acc. Chem. Res. 41 (2008) 157.
- [16] H. Valdés, V. Klusák, M. Pitoňák, O. Exner, I. Starý, P. Hobza, L. Rulíšek, J. Comput. Chem. 29 (2007) 861.
- [17] F. Neese, WIREs Comput. Mol. Sci. 2 (2012) 73.
- [18] S. Kozuch, J.M.L. Martin, Phys. Chem. Chem. Phys. 13 (2011) 20104.
- [19] S. Kozuch, D. Gruzman, J.M.L. Martin, J. Phys. Chem. C. 114 (2010) 20801.
- [20] DFT-D3 - A dispersion correction for density functionals, Hartree-Fock and semi-empirical quantum chemical methods, <http://www.thch.uni-bonn.de/tc/index.php?section=downloads&subsection=DFT-D3&lang=english>.
- [21] S. Kozuch, J.M.L. Martin, J. Comput. Chem. 34 (2013) 2327.
- [22] Y.K. Kang, H.S. Park, Chem. Phys. Lett. 600 (2014) 112.
- [23] M. Mališ, Y. Loquais, E. Gloaguen, H.S. Biswal, F. Piuze, B. Tardivel, V. Brenner, M. Broquier, C. Jouvet, M. Mons, N. Došlić, I. Ljubić, J. Am. Chem. Soc. 134 (2012) 20340.