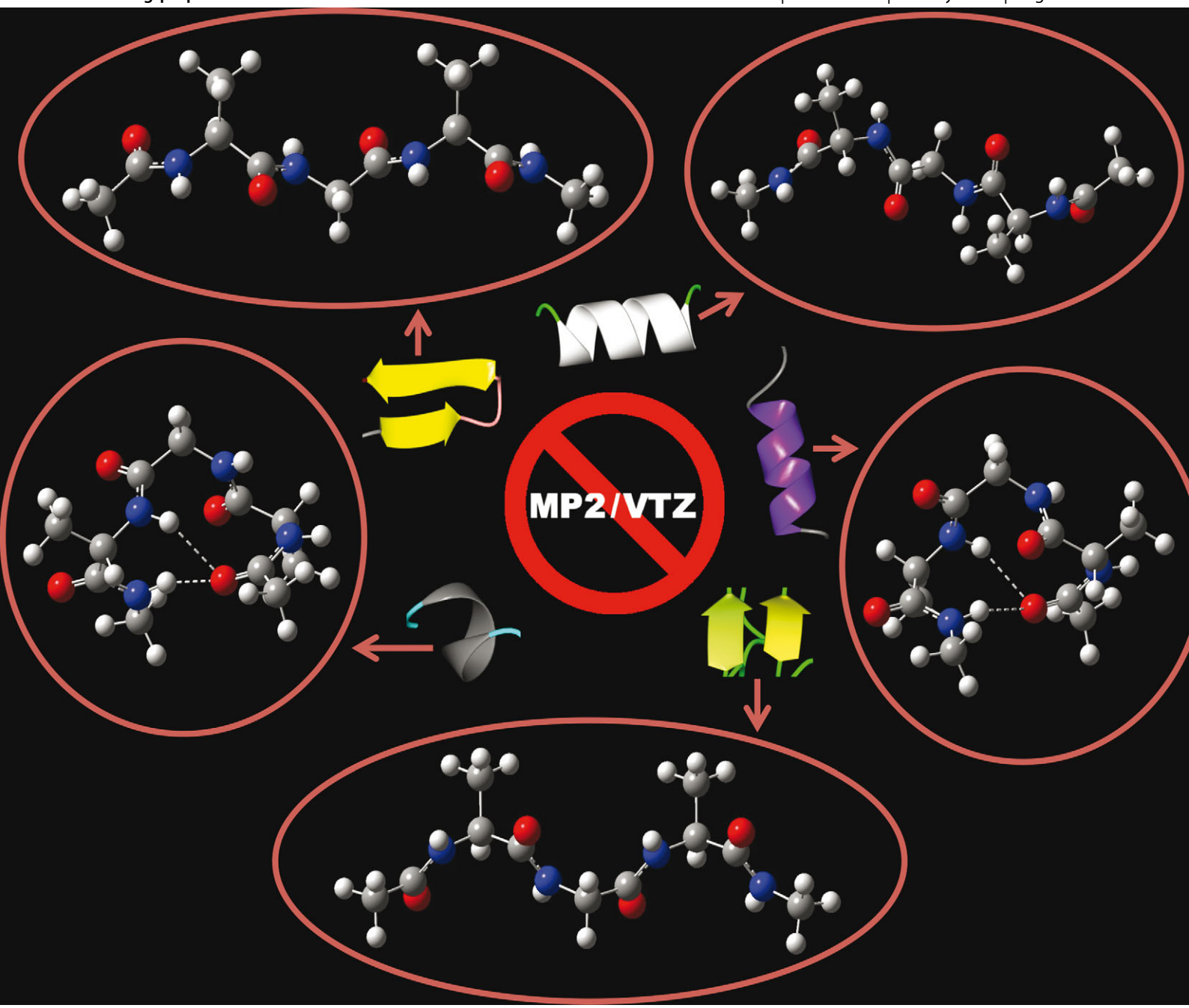
PCCP

Physical Chemistry Chemical Physics

www.rsc.org/pccp

Volume 15 | Number 19 | 21 May 2013 | Pages 7001–7422



ISSN 1463-9076

COMMUNICATION

Radom *et al*.

Accurate quantum chemical energies for tetrapeptide conformations: why MP2 data with an insufficient basis set should be handled with caution

1463-9076(2013)15:19;1-M

PCCP

RSC Publishing

COMMUNICATION

View Article Online
View Journal | View Issue

Open Access Article. Published on 13 February 2013. Downloaded on 2/10/2026 11:35:54 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Cite this: *Phys. Chem. Chem. Phys.*, 2013, **15**, 7028

Received 7th January 2013,
Accepted 29th January 2013

DOI: 10.1039/c3cp00057e

www.rsc.org/pccp

Accurate quantum chemical energies for tetrapeptide conformations: why MP2 data with an insufficient basis set should be handled with caution†

Lars Goerigk,‡*[a] Amir Karton,‡§*[ab] Jan M. L. Martin¶*[c] and Leo Radom*[ab]

High-level quantum chemical calculations have been carried out for biologically-relevant conformers of tetrapeptides. Our results indicate potential problems if the widely-applied MP2 approach is used in such situations with basis sets of insufficient size. Efficient alternatives are discussed.

Understanding the structure–function relationship in polypeptides and proteins is a crucial step in the elucidation of biochemical processes. In this regard, constantly improving hardware architectures and program codes have allowed the *in silico* treatment of proteins to become an important partner in related investigations. Due to the immense size of proteins, their computational study has generally been carried out with molecular mechanics (MM) force-field methods, such as CHARMM or GROMOS.[1] At the other extreme, high-level quantum mechanical (QM) methods, such as coupled-cluster with perturbative triples[2] [CCSD(T)] or composite methods like G*n*[3] or W*n*,[4] provide results with much higher accuracy. However, due to their computational cost, their applicability has been restricted to smaller systems such as amino acids, and di- and tripeptides.[5,6]

Tetrapeptides are the smallest model systems that are able to mimic the typical hydrogen-bond pattern in α-helices. They are therefore of particular biological interest, and QM treatments of them have been reported since the late 1990s.[7] The highest levels of theory used in such studies were often conventional or local second-order Møller–Plesset perturbation theory (MP2) with a triple-ζ basis set. These levels are still popular in recent similar investigations.[8] Of particular relevance to the present work is the extensive study of 100 tetrapeptides by Jiang *et al.*,[9] who evaluated the performance of a large number of QM and MM methods based on MP2/cc-pVTZ reference values.

In the present article, we will demonstrate where potential problems of the MP2 approach combined with finite basis sets might lie when applied to conformers of polypeptides. For this purpose, we selected two systems from the study by Jiang *et al.*[9] that have the sequence ACE-ALA-X-ALA-NME. ALA is alanine, X is either glycine (GLY) or serine (SER), and ACE and NME stand for acetyl and methylamide groups, respectively (see Fig. 1a). For each peptide, five conformers have been examined, whose backbone dihedral angles (see ESI†) have been fixed such that they resemble those typically found in parallel (β) and antiparallel (β_a) β-sheets, in right-handed (α_R) and left-handed (α_L) α-helices, and in the common polyproline-II (PP-II) helix. We note that while the two α-helical conformers are stabilised by hydrogen bonds, the backbones of the other three conformers are not (see for example Fig. 1).

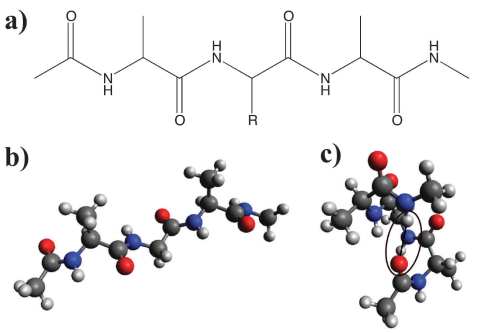


Fig. 1 (a) Schematic structure of the two tetrapeptide models with R = H or OH. (b) Structure of the β_a conformer for R = H. (c) Structure of the α_R conformer for R = H. The highlighted part shows the hydrogen bond.

[a] *School of Chemistry, University of Sydney, Sydney, NSW 2006, Australia. E-mail: lars.goerigk@chem.usyd.edu.au, radom@chem.usyd.edu.au; Fax: +61 2-9351-3329; Tel: +61 2-9351-2733*
[b] *ARC Centre of Excellence for Free Radical Chemistry and Biotechnology*
[c] *Department of Chemistry and Center for Advanced Scientific Computing and Modeling (CASCaM), University of North Texas, Denton, TX 76201, USA*

† Electronic supplementary information (ESI) available: Technical details, statistical data and relative and absolute energies for all tested methods, and structures of all tetrapeptides. See DOI: 10.1039/c3cp00057e

‡ L. Goerigk and A. Karton are equal contributors.

§ Current address: School of Chemistry and Biochemistry, University of Western Australia, Perth, Crawley, WA 6009, Australia. E-mail: amir.karton@uwa.edu.au; Fax: +61 8-6488-7330; Tel: +61 8-6488-3139

¶ Current address: Department of Organic Chemistry, Weizmann Institute of Science, IL-76100 Rehovot, Israel. E-mail: gershom@weizmann.ac.il; Fax: +972 8-934-4142; Tel: +972 8-934-2533

This journal is © the Owner Societies 2013

Open Access Article. Published on 13 February 2013. Downloaded on 2/10/2026 11:35:54 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

We present results for these biologically interesting tetrapeptide conformations based on a CCSD(T) approach extrapolated to the complete basis-set (CBS) limit. Due to the high cost of CCSD(T) for systems of this size, we estimated the 'true' CBS limit by using an additivity approach, the specific form of which has been shown to provide an accurate and reliable means for describing noncovalent interactions.[5,10–12] MP2/CBS results were obtained first and these values were then corrected for the difference between CCSD(T) and MP2 based on a small basis set (see ESI† for more details). The CCSD(T)/cc-pVDZ[13] calculations were the computationally most demanding and in total they took about 8.3 CPU years on Intel Nehalem 8837 cores (at 2.67 GHz) with 256 GB of RAM and 4 TB of disk.[14]

Our high-level relative conformational energies for the two peptides are shown in Table 1. For the glycine-containing system, the benchmark CCSD(T) ordering of the energies of its conformers is β_a, followed by the right-handed α_R-helix, the PP-II conformer, the α_L-helix and finally the β conformer. Our benchmark results allow an evaluation of less computationally demanding levels of theory (Table 1 and Fig. 2). Examination of the basis-set sequence from VDZ to VQZ to CBS for MP2 shows three categories of behaviour. The first category includes the β_a and β conformers. Their energy difference is well described by all levels of theory, reflecting their similar structural characteristics. The helical PP-II conformer belongs to the second category. Only a small basis-set dependence is observed and, except for VDZ, MP2 agrees reasonably well with the benchmark value. The third category comprises the two α-helices. For these, we observe a strong basis-set dependence of the relative conformational energies. At the VDZ, VTZ and VQZ levels, these two conformers are incorrectly predicted to be the most stable ones, *e.g.*, α_R is lower in energy than β_a by 7.3 kJ mol^{-1} for VTZ. The errors are reduced with increasing basis-set size. At the MP2/CBS level, β_a and α_R have virtually the same energies, followed by PP-II and α_L, whose energies are also close to one another.

Table 1 Calculated conformational energies (kJ mol^{-1})[a] obtained by various quantum chemical methods for the two tetrapeptides

Method	β_a	α_R	PP-II	α_L	β
ACE-ALA-GLY-ALA-NME					
CCSD(T)/CBS	0.0	2.4	4.4	8.0	8.5
CCSD(T)/VDZ	0.0	−13.1	−1.0	−11.2	7.5
MP2/VDZ	0.0	−15.1	1.4	−12.1	7.6
MP2/VTZ	0.0	−7.3	4.7	−0.8	8.1
MP2-F12/VDZ-F12	0.0	−0.1	7.1	6.7	8.4
MP2/VQZ	0.0	−2.2	6.2	4.4	8.7
MP2/CBS	0.0	0.4	6.9	7.1	8.6
PWPB95-D3/VQZ	0.0	4.5	8.4	9.1	8.9
DSD-PBEP86-D3/VQZ	0.0	2.3	6.3	8.6	8.4
ACE-ALA-SER-ALA-NME					
CCSD(T)/CBS	0.0	4.4	11.0	7.5	11.1
CCSD(T)/VDZ	0.0	−10.7	3.5	−11.4	9.9
MP2/VDZ	0.0	−12.9	5.6	−11.6	10.5
MP2-F12/VDZ-F12	0.0	1.7	13.4	6.9	11.5
MP2/VTZ	0.0	−4.7	10.4	0.8	11.6
MP2/VQZ	0.0	−0.2	12.0	5.1	11.8
MP2/CBS	0.0	2.2	13.1	7.3	11.7
PWPB95-D3/VQZ	0.0	6.0	15.0	7.6	12.2
DSD-PBEP86-D3/VQZ	0.0	4.2	12.4	8.8	16.1

[a] Energies calculated relative to the β_a conformer.

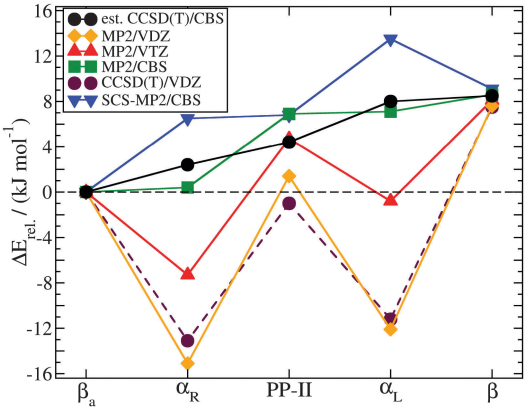


Fig. 2 Energies given by six different methods for four tetrapeptide conformers relative to the β_a conformer for the glycine-containing tetrapeptide.

Our results suggest that the most likely source for the basis-set dependence of the third category is an overstabilisation of the hydrogen bonds for finite basis sets. This is likely to be associated with an intramolecular basis-set-superposition error. Similar behaviour is observed for CCSD(T)/VDZ, indicating that the problem is largely caused by insufficient basis-set size and not inherent to MP2 itself.[15] Addition of diffuse functions has been reported to improve the description of hydrogen bonds.[15,16] However, in this case the improvement is found to be small (see ESI†), and it is necessary to go closer to the CBS limit to treat hydrogen bonds adequately. Overall, the performance of MP2/CBS is very good, which arises because the added CCSD(T) correction is small for the tetrapeptides that we have examined.

One alternative to conventional MP2 is the spin-component-scaled MP2 (SCS-MP2) method,[17] which has been shown to be often more accurate and robust than MP2.[6,17] However, it is also known that it underestimates the strength of hydrogen bonds,[18] as we clearly see for SCS-MP2/CBS for the α-helices (Fig. 2).

Additional alternatives to conventional MP2 are approaches that explicitly take into account the interelectronic distances in the wave function. We investigate here the MP2-F12 method,[19] which achieves a faster convergence to the CBS limit than conventional MP2. For the present systems, this can already be seen at the double-ζ level: MP2-F12/VDZ-F12[20] yields relative energies comparable to MP2/CBS (see Table 1). MP2-F12/VTZ-F12 results are almost identical to both MP2-F12/VDZ-F12 and MP2/CBS (see ESI†).

The conformational preferences for the serine-containing peptide are similar to those for the glycine system (Table 1). The major difference is the significantly higher relative energy of PP-II, which becomes energetically similar to the β conformer.

Density functional theory (DFT) approximations have become the 'work-horse' of quantum chemistry and it is therefore interesting to also discuss their performance for the present systems. We tested various density functionals that have been shown previously to be accurate for related cases (see ESI†).[6] These DFT calculations were carried out with the

Open Access Article. Published on 13 February 2013. Downloaded on 2/10/2026 11:35:54 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

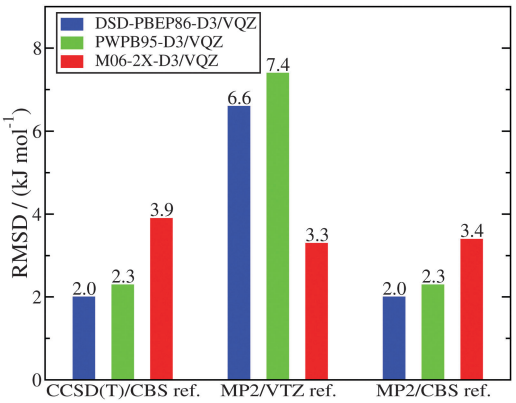


Fig. 3 Root-mean-square deviations (RMSDs) of conformational energies for three dispersion-corrected DFT procedures with respect to three different sets of reference values for the two tetrapeptides.

VQZ basis set and Grimme's London-dispersion correction DFT-D3.[21] In accordance with previous work, we find the most promising DFT methods to be double-hybrid functionals,[22] which represent a combination of a standard density functional with portions of Hartree–Fock exchange and (SCS-)MP2 correlation. Results for the best-performing double-hybrids, DSD-PBEP86-D3[23] and PWPB95-D3,[24] are included in Table 1. With a few exceptions, the correct trends of the conformer ordering are predicted for both systems. However, resolving the relative energies of α_L and β structures seems to be problematic for the glycine system for both methods, the energy of PP-II is overestimated for both peptides by PWPB95-D3, and DSD-PBEP86-D3 overestimates the energy difference between β_a and β in the serine case. Considering that these functionals are among the best currently available, our results show that the tetrapeptide conformers represent a difficult case for current DFT methods.

Our assessment of DFT procedures used the estimated CCSD(T)/CBS energies as a benchmark. However, in related previous studies, MP2 energies obtained with finite basis sets have been used as a reference for the evaluation or parameterization of computationally less demanding methods.[7–9] We demonstrate the importance of the quality of the reference values in Fig. 3, which shows root-mean-square deviations (RMSDs) for DSD-PBEP86-D3, PWPB95-D3, and M06-2X-D3[25] for the full set of relative energies for the two tetrapeptides. Three sets of reference values are considered: CCSD(T)/CBS, MP2/VTZ, and MP2/CBS. Compared with the CCSD(T)/CBS values, the two double-hybrids have the lowest and M06-2X-D3 the highest RMSDs of the three methods. This picture changes completely when MP2/VTZ is used as the benchmark. M06-2X-D3 would then be interpreted as the most suitable, while the double-hybrid functionals would be considered to be significantly worse. On the other hand, the trends based on an MP2/CBS reference closely parallel those for the CCSD(T)/CBS case.

In summary, our estimated CCSD(T)/CBS relative energies of conformers of two biologically relevant tetrapeptides have allowed an evaluation of the widely-used MP2 approach. We conclude that it is desirable for conventional MP2 calculations of conformational energies to be carried out with large basis sets or at the CBS limit to obtain a balanced description of conformers with and without hydrogen bonds. An efficient way to achieve such results is through the MP2-F12 approach in combination with a small basis set, which yields relative energies very close to the MP2/CBS limit but at a significantly lower computational cost. Provided that the CCSD(T)–MP2 correction is small, MP2/CBS is also close to the CCSD(T)/CBS limit. Dispersion-corrected double-hybrid DFT turns out to be the currently best DFT approach. It provides a useful alternative to conventional MP2[6] but has some shortcomings for the present systems. We hope that our estimated CCSD(T)/CBS energies will be useful for future method development.

We gratefully acknowledge funding to L.G. from the German Academy of Sciences "Leopoldina" Fellowship Programme (grant number LPDS 2011-11), and to A.K. and L.R. from the Australian Research Council, and generous allocations of computer time from CASCaM, from the NCI National Facility, and from Intersect Australia Ltd.

Notes and references

1 X. Zhu, P. E. M. Lopes and A. D. MacKerell, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2012, **2**, 167; C. Oostenbrink, A. Villa, A. E. Mark and W. F. van Gunsteren, *J. Comput. Chem.*, 2004, **25**, 1656.
2 K. Raghavachari, G. W. Trucks, J. A. Pople and M. Head-Gordon, *Chem. Phys. Lett.*, 1989, **157**, 479.
3 L. A. Curtiss, P. C. Redfern and K. Raghavachari, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2011, **1**, 810.
4 A. Karton, S. Daon and J. M. L. Martin, *Chem. Phys. Lett.*, 2011, **510**, 165; A. Karton and J. M. L. Martin, *J. Chem. Phys.*, 2012, **136**, 124114.
5 D. Řeha, H. Valdes, J. Vondrasek, P. Hobza, A. Abu-Riziq, B. Crews and M. S. de Vries, *Chem.–Eur. J.*, 2005, **11**, 6803; H. Valdes, K. Pluháčková, M. Pitonák, J. Řezáč and P. Hobza, *Phys. Chem. Chem. Phys.*, 2008, **10**, 2747.
6 L. Goerigk and S. Grimme, *Phys. Chem. Chem. Phys.*, 2011, **13**, 6670.
7 M. D. Beachy, D. Chasman, R. B. Murphy, T. A. Halgren and R. A. Friesner, *J. Am. Chem. Soc.*, 1997, **119**, 5908; N. Gresh, S. A. Kafafi, J.-F. Truchon and D. R. Salahub, *J. Comput. Chem.*, 2004, **25**, 823.
8 C. Steinmann, D. G. Fedorov and J. H. Jensen, *PLoS One*, 2012, **7**, e41117; A. Buczek, R. Walesa and M. A. Broda, *Biopolymers*, 2012, **97**, 518; M. A. Alvarez, E. J. Saavedra, M. S. Olivella, F. D. Suvire, M. A. Zamoraa and R. D. Enriz, *Cent. Eur. J. Chem.*, 2012, **10**, 248.
9 J. Jiang, Y. Wu, Z.-X. Wang and C. Wu, *J. Chem. Theory Comput.*, 2010, **6**, 1199.
10 P. Jurečka and P. Hobza, *Chem. Phys. Lett.*, 2002, **365**, 89.
11 We have shown the applicability of this approach to general thermochemistry in a separate investigation.
12 For example, see also: J. M. L. Martin, *Theor. Chem. Acc.*, 1997, **97**, 227; W. Klopper and H. P. Lüthi, *Mol. Phys.*, 1999, **96**, 559; D. G. Liakos and F. Neese, *J. Phys. Chem. A*, 2012, **116**, 4801.

 This journal is © the Owner Societies 2013

Open Access Article. Published on 13 February 2013. Downloaded on 2/10/2026 11:35:54 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

13 T. H. Dunning, Jr., *J. Chem. Phys.*, 1989, **90**, 1007.
14 We use the acronyms V*n*Z for the cc-pV*n*Z basis sets.
15 A. Halkier, H. Koch, P. Jørgensen, O. Christiansen, I. M. Beck Nielsen and T. Helgaker, *Theor. Chem. Acc.*, 1997, **97**, 150.
16 P. Jurečka, J. Černy, P. Hobza and D. Salahub, *J. Comput. Chem.*, 2007, **28**, 555.
17 S. Grimme, *J. Chem. Phys.*, 2003, **118**, 9095.
18 J. Antony and S. Grimme, *J. Phys. Chem. A*, 2007, **111**, 4862; R. A. Bachorz, F. A. Bischoff, S. Höfener, W. Klopper, P. Ottiger, R. Leist, J. A. Frey and S. Leutwyler, *Phys. Chem. Chem. Phys.*, 2008, **10**, 2758; A. Karton, R. J. O'Reilly, B. Chan and L. Radom, *J. Chem. Theory Comput.*, 2012, **8**, 3128; S. Grimme, L. Goerigk and R. F. Fink, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2012, **2**, 886.
19 H.-J. Werner, T. B. Adler and F. R. Manby, *J. Chem. Phys.*, 2007, **126**, 164102.
20 K. A. Peterson, T. B. Adler and H.-J. Werner, *J. Chem. Phys.*, 2008, **128**, 084102.
21 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104; S. Grimme, S. Ehrlich and L. Goerigk, *J. Comput. Chem.*, 2011, **32**, 1456.
22 S. Grimme, *J. Chem. Phys.*, 2006, **124**, 034108.
23 S. Kozuch and J. M. L. Martin, *Phys. Chem. Chem. Phys.*, 2011, **13**, 20104.
24 L. Goerigk and S. Grimme, *J. Chem. Theory Comput.*, 2011, **7**, 291.
25 Y. Zhao and D. G. Truhlar, *Theor. Chem. Acc.*, 2008, **120**, 215.