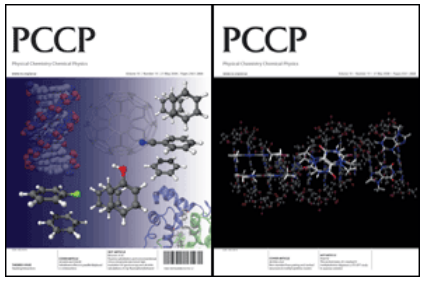
PCCP
Physical Chemistry Chemical Physics



This paper is published as part of a PCCP Themed Issue on:

Stacking Interactions

Guest Editor: Pavel Hobza

Editorial

Perspectives

Papers

Benchmark database on isolated small peptides containing an aromatic side chain: comparison between wave function and density functional theory methods and empirical force field†

Haydee Valdes,* Kristýna Pluháčková, Michal Pitonák, Jan Řezáč and Pavel Hobza*

Received 14th December 2007, Accepted 18th February 2008
First published as an Advance Article on the web 13th March 2008
DOI: 10.1039/b719294k

A detailed quantum chemical study on five peptides (WG, WGG, FGG, GGF and GFA) containing the residues phenylalanyl (F), glycyl (G), tryptophyl (W) and alanyl (A)—where F and W are of aromatic character—is presented. When investigating isolated small peptides, the dispersion interaction is the dominant attractive force in the peptide backbone–aromatic side chain intramolecular interaction. Consequently, an accurate theoretical study of these systems requires the use of a methodology covering properly the London dispersion forces. For this reason we have assessed the performance of the MP2, SCS-MP2, MP3, TPSS-D, PBE-D, M06-2X, BH&H, TPSS, B3LYP, tight-binding DFT-D methods and ff99 empirical force field compared to CCSD(T)/complete basis set (CBS) limit benchmark data. All the DFT techniques with a '-D' symbol have been augmented by empirical dispersion energy while the M06-2X functional was parameterized to cover the London dispersion energy. For the systems here studied we have concluded that the use of the ff99 force field is not recommended mainly due to problems concerning the assignment of reliable atomic charges. Tight-binding DFT-D is efficient as a screening tool providing reliable geometries. Among the DFT functionals, the M06-2X and TPSS-D show the best performance what is explained by the fact that both procedures cover the dispersion energy. The B3LYP and TPSS functionals—not covering this energy—fail systematically. Both, electronic energies and geometries obtained by means of the wave-function theory methods compare satisfactorily with the CCSD(T)/CBS benchmark data.

Introduction

Peptides are relevant biological systems either as isolated molecules or as protein building blocks and for this reason they have received prominence in molecular biology. Typically, the studies performed on peptides deal with these macromolecular structures in solution. Valuable complementary information can be also obtained from their study in the gas phase. Gas-phase studies are the gate to the chemical and physico-chemical properties of any molecule and consequently, all the intrinsic properties of the peptides can be obtained without any risk of them being washed out or masked by the environment. The most important property of a peptide is its structure and evidently, gas phase studies represent the first step in determining the structure in an environment. This is an extremely difficult task since here not only covalent but also non-covalent interactions play a key role. Evaluation of non-covalent interactions in such extended systems like peptides is difficult and represents one of the most challenging tasks of today's computational chemistry. The main reason is that peptides are very flexible systems showing an extensive conformational landscape. In an ideal study, nonempirical *ab initio* molecular dynamics simulations should be performed so that each of the different conformers of the peptide would be accurately calculated. But this procedure is still far from being practical. Thus, nowadays philosophy is to use a more *modest* method based on lower-level potentials (mostly empirical potentials) to generate the conformational landscape of the peptide and next, select only few of the most stable conformers for a more accurate quantum chemical treatment. Since the existing experiments are mostly done at room temperature it is necessary to pass from the potential energy surface (PES) to the free energy surface (FES). Let us add here that the evaluation of the entropy is difficult and reliable calculations of the FES are more demanding than in the PES.

A well established procedure to explore the FES of a biomolecule is to combine the molecular dynamics/quenching (MD/Q) technique[1] with an empirical potential. Applying the MD/Q technique provides all the energy minima of the surface as well as the population of each individual minimum, which in turn is proportional to its free energy. However, we have recently proven that MD/Q simulations performed with the AMBER (ff99) force field provide inaccurate results for the

Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic and Center for Biomolecules and Complex Molecular Systems, 166 10 Prague 6, Czech Republic. E-mail: haydee.valdes@marge.uochb.cas.cz; pavel.hobza@uochb.cas.cz; Fax: +420-220 410 320

† Electronic supplementary information (ESI) available: Fig. S1, Table S1 and RI-MP2/cc-pVTZ Cartesian coordinates of all the conformers included in the database. See DOI: 10.1039/b719294k

study of isolated peptides containing aromatic side chains. There are various reasons for it but probably the most important one is the fact that the empirical potential works with effective average charges (over all existing structures) which might be far from the charges describing properly each individual structure.[2,3] This inconvenient can be overcome by performing MD/Q simulations in combination with the tight-binding density functional theory (DFT) method in its augmented version by an empirical dispersion term (DF-TB-D).[4] Then, the problem mentioned above is here automatically removed since electron densities are determined for each structure of the system. Despite the fact that the tight-binding DFT-D is very fast (by several orders of magnitude in comparison with the standard DFT techniques) this procedure is computationally more expensive than when an empirical force field is used and consequently, the length of the MD/Q trajectory has to be reduced. This means that the MD/Q technique in combination with tight-binding DFT-D is not an efficient tool for scanning the FES of an isolated peptide since here much longer simulation times are required than in the case of scanning the PES. Yet, it can be very well used for the localization of all the minima in the PES. Besides, in combination with standard statistical thermodynamic calculations based on the rigid rotor-harmonic oscillator-ideal gas approximation (RR-HO-IG) using the *ab initio* constants, it constitutes a reasonable bypass to the FES. Thus, the combination of MD/Q simulations performed using the tight-binding DFT-D method with high-level correlated *ab initio* quantum chemical calculations including the RR-HO-IG approximation for determining the thermodynamic characteristics (the abbreviation MD/Q + QM is used throughout the text) has been the procedure used all over this work in order to explore the FES of the peptides here studied.

Using the MD/Q + QM procedure has one serious drawback, namely the description of the systems is harmonic. An elegant solution to the problem is to perform on-the-fly MD simulations covering the anharmonic effects. Currently, we are performing on-the-fly MD simulations for the phenylalanyl-glycyl-glycine (FGG) tripeptide for which we have infrared (IR) experimental data.[2] Preliminary results show that the theoretical harmonic and anharmonic far IR spectra are similar and agree well with the experimental results. Regarding the mid-IR spectrum, it seems that on-the-fly MD simulations spectra have to be calculated at a relatively high level of theory in order to reproduce the experimental data. Obviously then, the applicability of this methodology is currently inhibited by its computational requirements. We have also applied the metadynamics free energy modeling technique[5,6] for the study of the FES of an isolated peptide.[3] This method offers two major advantages over the MD/Q + QM procedure namely, (a) it allows exploring the FES at the tight-binding DFT-D level of theory at a very reasonable computational cost and (b) it includes the anharmonic effects. The results obtained from the metadynamic calculations agree well with those obtained from the MD/Q + QM procedure as well as with the experimental data, reinforcing thus, the suitability of the latter, for the study of isolated small peptides. Notice that, a free energy surface calculated by metadynamics is a function of a limited number of collective variables (typically two) and obviously, the information revealed in the plotted FES depends very much upon the selection of variables made. As a consequence, relevant data can be hidden.

The quantum chemical information obtained on the electronic structures and energies—along with many other properties—as well as on the existing non covalent interactions is certainly very interesting for its biological relevance. Notice for instance that the structural data on peptides can be very helpful in understanding the secondary structure of proteins.[7–9] But peptides are also very interesting from a purely quantum chemical point of view as multiconformational systems showing plenty of local minima within a very small range of energy. Indeed, the final order of structures is strongly affected by the level of theory employed for the *ab initio* calculations. For this reason, we present here a benchmark database of relative energies and geometries for a set of 76 conformers of different isolated small peptides containing an aromatic side chain calculated at the CCSD(T) complete basis set limit (CBS) level of theory. The peptides chosen are the following: tryptophyl–glycine (WG) dipeptide, phenylalanyl–glycyl–glycine (FGG), glycyl–glycyl–phenylalanine (GGF), glycyl–phenylalanyl–alanine (GFA) and tryptophyl–glycyl–glycine (WGG) tripeptides (see the following section for a detailed description on the selection of the database structures).

Up to our knowledge, no database built out from isolated systems of biological relevance has been published yet and therefore, we believe that the one here provided can be very useful for the further assessment of newly developed methodologies as well as for the improvement of the nowadays available empirical force fields. This last point is of particular relevance since experimental data on neutral small peptides in the gas-phase are relatively sparse. The main reason is that rotational spectroscopic techniques provide spectra which are very difficult to unravel for larger systems than dipeptides and, vibrational spectroscopic techniques do not provide enough structural information of the system under study. Additionally, a vast majority of the experimental studies are limited to the far infrared region in the interval between 3000 and 4000 cm^{-1}.

A final comment should be done on the importance of the dispersion energy for the stabilization of the peptides here studied. The role of dispersion energy and its proper theoretical description is a matter of concern in the theoretical study of any molecule—and particularly in biomolecules—and it constitutes one of the issues most thoroughly investigated in our laboratory first, in molecular complexes and now also, in isolated molecules such as small peptides. We have recently proven that the London dispersion forces are the major attractive forces in the peptide backbone–aromatic side chain intramolecular interaction.[3] The proper treatment of the dispersion energy requires coupled cluster calculations including non-iterative triple excitations and the use of a medium-size basis set. However, due to the system size—overall when thinking about the study of larger biomolecules than tripeptides—such calculations are not feasible. Then, it would be desirable to have at our disposal a method combining reasonable computational requirements with a proper description of the dispersion energy. Aiming to find an alternative level of

theory to CCSD(T)/*medium size basis set* for the study of isolated peptides we have explored various possibilities, namely: (a) the fast ff99 empirical force field;[10] (b) the semi-empirical tight-binding DFT method augmented with an empirical term to include the dispersion energy;[4] (c) functionals which include the dispersion term directly (TPSS-D,[11] PBE-D[12–14]) or indirectly *via* parameterization (M06-2X[15]) or standard functionals (BH&H,[16] TPSS[17] and B3LYP[18]) and, (d) second (MP2) and third (MP3) order Møller–Plesset[19] and spin-component scaled second-order Møller–Plesset (SCS-MP2)[20] wave function theory (WFT) methods. The choice of these WFT methods requires a comment. On one hand, the MP2 method is known to overestimate the dispersion energy when an extended basis set (or even at the CBS limit) is applied. Reliable energies (and also geometries) are thus frequently obtained when using medium size basis sets. However, this is evidently due to a compensation of errors and it is impossible to rely on this compensation. On the other hand, the use of the MP3 and SCS-MP2 procedures is attractive since these methods reduce the overestimation of the dispersion energy. As for the basis sets, basis sets of the Pople's,[21] Truhlar's[22] and Dunning's[23] laboratories have been here employed. The latter have been further used to carry out extrapolations to the complete basis set limit not only for the sake of obtaining accurate energies and geometries, but also to correct for the intramolecular basis set superposition error.[24–26] It must be mentioned here that only systematically improved basis sets can be used for extrapolation purposes.

Finally, we would like to mention that peptides containing aromatic side chains share a significant feature with intermolecular aromatic complexes and this is the importance of the stacking interactions. For this reason we decided to include the present study in the special issue of the *Phys. Chem. Chem. Phys.* journal devoted to stacking.

Methods and computational details

Selection of structures for the database

Each peptide has been systematically studied following the strategy of calculation described in Chart 1. Essentially, MD/Q simulations performed with the tight-binding DFT-D method are combined with high-level correlated *ab initio* quantum chemical calculations. By means of the MD/Q simulations we sample the conformational landscape of each peptide in order to localize all the possible existing minima. Then, the most stable ones are selected in consecutive steps according to their relative energies calculated at different levels of theory. Our philosophy is to decrease the number of structures considered as we increase the level of theory employed. The final output is a set containing the 15 most stable conformers in the PES of each peptide. Notice that the number of structures included in the final set is somehow arbitrary. However, based on our previous experience in isolated small peptides,[2,3,27] we believe that this constitutes a reasonable cutoff since on one hand we can be sure that the set includes those structures observed experimentally and on the other hand, the number of calculations to be performed for a benchmark study is still *easy* to handle. Consequently, the benchmark database here presented contains 76 structures.

Molecular dynamics/quenching (MD/Q) technique

The PES of each individual peptide here studied has been scanned by means of the self-consistent charge density functional tight-binding method extended by an empirical dispersion term (SCC-DF-TB-D)[4] using the molecular dynamics/quenching (MD/Q) technique.[1] Essentially, each MD simulation—performed with scaled velocities at a temperature of 900 K—was repeatedly quenched after a limited number of steps. Then the kinetic energy term was removed and a non restricted minimization using the conjugate gradient method was performed. Once the output energies and coordinates of the resulting minima were stored, the MD simulation was continued from the same point before the quench was done. Both, the MD and Q were done using the SCC-DF-TB-D method. In the tight-binding DFT method,[28] the equations applied are derived from a second-order expansion of the DFT total energy functional with respect to charge density fluctuations about a given reference density. Subsequently, the method was improved by adding an empirical term covering the London dispersion forces.[4] Thus, by definition, tight-binding DFT-D is superior to any of the standardly-used empirical force fields. Additionally, as mentioned before, probably the most important advance of the method concerns the fact that it does not require any specification of atomic charges which might be very different for various geometrically distinct conformers.

The screening of the PESs is followed by a depuration step. The reason is that we may localize repeatedly the same minima during the exploration of the PESs. Consequently, the MD/Q output structures have to be energetically ordered and geometrically distinguished. The final set of minima obtained will be further used as an input for the *ab initio* study.

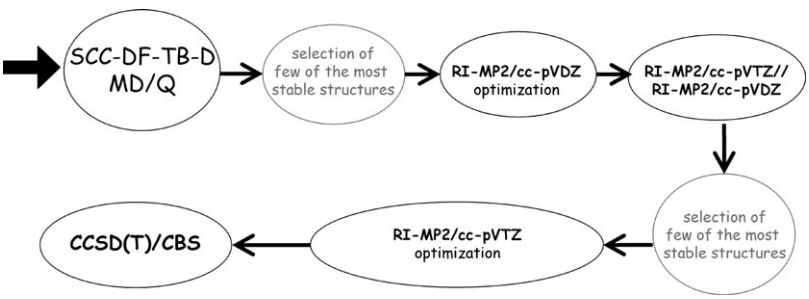


Chart 1 Strategy of calculation followed for the construction of the benchmark database.

The SCC-DF-TB-D method has been also used to carry out single-point energy calculations as well as geometry optimizations of the whole set of structures included in the database.

Wave function theory (WFT) calculations

Benchmark geometries were optimized at the RI[29,30]-MP2[19]/cc-pVTZ[23] level of theory whereas benchmark relative energies were obtained using the CCSD(T)[31,32] method according to the following equation:

$$E^{CCSD(T)}{}_{CBS} = E^{MP2}_{CBS} + (E^{CCSD(T)} - E^{MP2})|_{6\text{-}31G^*} \quad (1)$$

where E^{MP2}_{CBS} is obtained from the extrapolation to the complete basis set (CBS) limit of the RI-MP2/cc-pVTZ and RI-MP2/cc-pVQZ relative energies according to the scheme of Helgaker and co-workers.[33] The second term describes the higher-order contributions to the correlation energy (beyond the second perturbation order) and this energy difference depends only negligible on the quality of the basis set.[34] This is the reason why only a rather small basis set (6-31G*) was applied. Essentially the same equation has been applied in order to obtain the MP3/CBS values. The only difference being that the second term is expressed as $(E^{MP3} - E^{MP2})|_{6\text{-}31G^*}$.

Finally, we have recalculated the relative energies of the different structures by means of the spin-component scaled second-order Møller–Plesset (SCS-MP2) method,[20] with 1.2 and 1/3 scaling factors for antiparallel and parallel spin, respectively. Additionally, the SCS-MP2 calculations were accelerated with the density fitting program as implemented in Molpro quantum chemistry package. SCS-MP2 relative energies were calculated using both, cc-pVTZ and cc-pVQZ Dunning's basis sets and later extrapolated to the CBS limit using the same scheme of extrapolation than for the MP2 relative energies.

Density functional theory (DFT)

B3LYP functional. B3LYP[18]/6-311++G(3df,3pd)[21] (abbreviated as LP_1 in this paper) single-point energy calculations for the whole set of structures and B3LYP/6-31G*[21] geometry optimizations for the P26 trial set (see geometrical analysis for the definition of the P26 trial set) were carried out. B3LYP is a hybrid functional including a mixture of Hartree–Fock exchange with DFT exchange–correlation.

TPSS functional with and without an empirical dispersion term (TPSS-D/TPSS). TPSS[17]/LP_1 and TPSS-D[11]/LP_1 single-point energy calculations have been carried out for the 76 conformers included in the database. Besides, TPSS-D/LP_1 geometry optimizations were performed for the P26 trial set. The TPSS functional is a nonempirical meta-generalized gradient approximation functional. Its augmented version by dispersion energy, *i.e.* TPSS-D, has been recently proven to perform reasonably well for the study of noncovalent complexes[11,35] and isolated systems.[3,36] The London dispersion energy is included by a damped pair-potential parameterized against CCSD(T)/CBS results for model complexes containing H-bonded, dispersion-controlled and mixed complexes.[11]

PBE functional. The PBE[12–14] functional of Perdew, Burke and Ernzerhof has the following form:

$$0.75(S + PBE(X)) + 0.25HF + PW + PBE(C) \quad (2)$$

PBE(X) and PBE(C) are the Perdew–Burke–Ernzerhof exchange and correlation functionals[13,14] and PW is the Perdew–Wang correlation functional.[12] Like the TPSS functional, the PBE has been augmented by an empirical term (the same as the one used for TPSS-D; the parameters in the damping functions are, however, in these functionals different) covering the dispersion energy. Then, PBE-D/6-311++G(3df,3pd)[21] single-point energy calculations were carried out for the 76 conformers included in the database.

M06-2X functional. The M06-2X[15] functional of Truhlar and co-workers has been used in combination with the MIDI![22] and 6-311+G(2df,2p)[21] basis sets (abbreviated as LP_2 in this paper) for geometry optimizations and single-point energy calculations, respectively. The use of the smaller MIDI! basis set for geometry optimization was recommended in the original paper. The M06-2X functional belongs to a new generation of hybrid meta-generalized-gradient-approximation exchange–correlation functionals which include an accurate treatment of medium-range correlation energy what mainly concerns the London dispersion energy. This is done by parameterization of an extended set of parameters in the exchange part of the functional. This functional has been proven efficient for the prediction of noncovalent interactions.

Becke half-and-half functional (BH&H). Single-point energy calculations at the BH&H[16]/6-311++G(d,p)[21] (abbreviated as LP_3 in this paper) level of theory have been here performed for all the constituent structures of the database. The BH&H functional is a hybrid functional constructed as a sum of the Lee, Yang, and Parr's correlation functional and 50% of each: local density approximation and of exact Hartree–Fock exchange. It was shown recently that this potential covers some portion of dispersion energy and thus provides rather reliable results for molecular stacking. Simultaneously it was demonstrated that it is less efficient for clusters with H-bonds.[35]

ff99 empirical force field. Single-point energy calculations and geometry optimizations were carried out by means of ff99 empirical force field[10] using the parameters contained in the parm99.dat file. Two sets of restrained electrostatic potential fitting procedure[37] (RESP) atomic charges have been used, *i.e.* B3LYP/cc-pVTZ and HF/6-31G* RESP charges. Each set was obtained averaging the atomic charges of few geometrically different conformations.

Codes

All the calculations have been carried out with the following programs: (a) MD/Q simulations as well as geometry optimizations of all the structures of the database and their single-point energy calculations at the SCC-DF-TB-D level of theory were carried out with the DFTB+ program.[38] (b) CCSD(T), MP2 and SCS-MP2 single-point calculations were done using the MOLPRO 2002.1 program.[39] (c)MP3/CBS single-point calculations were done using the MOLCAS 7 program package employing Cholesky-decomposed two electron integrals.[40,41] (d) RI-MP2/cc-pVTZ geometry optimizations and RI-DFT-D (in our own implementation) geometry

optimizations (TPSS-D/LP_1 and B3LYP/6-31G*) and single-point calculations (TPSS-D/LP_1, PBE-D/LP_1 B3LYP/LP_1 and TPSS/LP_1) were carried out using the TURBOMOLE 5.9 program package.[42] (e) M06-2X/LP_2 single-point calculations and M06-2X/MIDI! geometry optimizations were performed using QChem 3.1.[43] (e) All the calculations using the ff99 force field were done with AMBER9 molecular dynamic package.[44] (f) BH&H/LP_3 single point calculations were performed using the Gaussian 03 program suite.[45] (g) We used our own scripts for the selection of the geometrically distinct structures.

Results and discussion

Geometrical analysis

The geometrical analysis has been performed on 26 out of the 76 structures considered for the benchmark relative energy calculations (see Table 1 and shady insets in Fig. S1 of the Supporting information† for the chosen structures). The selection of the different conformers has been made based on two criteria (see Fig. 1): (a) different conformational arrangements of the peptide backbone and (b) different backbone/aromatic side chain intramolecular interactions. More specifically, first we have taken conformations showing extended and folded backbone arrangements but disregarding structures from these subgroups that showed minor geometrical differences as, for instance, the rotation of the $-NH_2$ terminus group. Subsequently, we have selected structures showing different backbone/aromatic side chain intramolecular interactions within these two subgroups. We believe this database (named P26) could be further used as a trial set, for instance, for the assessment of newly developed methodologies. Let us add here that the same philosophy was applied in our nucleic acid database set published recently in the same journal[46] and thus, the present P26 set is to some extent analogous to the S22 set from the reference mentioned.

The overall conclusion of our previous works on peptides,[2,3,27,36] is the important role played by the dispersion energy in the stability of the different conformations of a peptide containing an aromatic side chain. Thus, the use of methodologies covering properly the London dispersion energy is essential for the theoretical study of these systems. Keeping this fact in mind we have performed the geometry optimizations at the RI-MP2/cc-pVTZ theory. Besides, we have also included in the analysis the B3LYP functional in combination with the 6-31G* basis set. The assessment of this functional is particularly topical since it is being systematically used by the scientific community in the study of isolated small peptides containing aromatic side chains. Additionally, geometries optimized using the ff99 force field and two different sets of RESP charges (HF/6-31G* and B3LYP/cc-pVTZ) are also included in the analysis. Finally it should be mentioned, that we have not performed optimizations using the BH&H functional since in a recent study on nucleic acid base trimers it was concluded that the BH&H functional gives geometries of comparable quality to other DFT functionals.[35] Since we have performed geometry optimizations using the TPSS-D functional, obviously we have not carried out any optimization using the TPSS functional. Finally since the performance of the PBE-D functional is not that satisfactory (see energetic analysis) we have neither reoptimized the P26 database with this functional.

Table 1 collects for the P26 database the root-mean-square deviations (RMSDs) in Å between the RI-MP2/cc-pVTZ benchmark geometries and the different levels of theory here examined. Average RMSDs values are also included. The first

Table 1 Root-mean-square deviations (RMSDs) in Å between the RI-MP2/cc-pVTZ benchmark geometries and different levels of theory. Average RMSDs values are also included. LP_1 stands for 6-311++G(3df,3pd) Pople basis set. ff99(HF) and ff99(DFT) refer to relative energies calculated with the ff99 force field using HF/6-31G* and B3LYP/cc-pVTZ sets of RESP charges, respectively

	TPSS-D/LP_1	M06-2X/MIDI!	B3LYP/6-31G*	SCC-DF-TB-D	FF99(DFT)	FF99(HF)
WG_03	0.11	0.18	1.21	0.09	0.14	0.18
WG_06	0.11	0.19	0.41	0.10	0.18	0.15
WG_10	0.09	0.24	1.47	0.04	0.16	0.16
WG_14	0.08	0.17	0.69	0.04	0.15	0.14
WG_15	0.07	0.12	0.20	0.05	0.22	0.18
WGG_03	0.15	0.46	0.31	0.07	0.12	0.12
WGG_04	0.09	0.10	0.19	0.06	0.15	0.17
WGG_05	0.08	0.25	0.28	0.26	0.17	0.21
WGG_06	0.18	0.12	1.18	0.06	0.11	0.15
WGG_08	0.30	0.15	0.42	0.08	0.17	0.13
WGG_11	0.11	0.18	0.94	0.28	0.19	0.19
FGG_055	0.09	0.18	0.18	0.04	0.11	0.12
FGG_099	0.07	0.16	0.27	0.09	0.13	0.16
FGG_215	0.64	0.12	0.63	0.05	0.11	0.12
FGG_252	0.16	0.29	0.34	0.15	0.14	0.28
FGG_357	0.07	0.10	0.50	0.05	0.24	0.17
FGG_412	0.31	0.19	0.68	0.05	0.12	0.14
FGG_470	0.08	0.18	0.51	0.05	0.12	0.13
GGF_01	0.08	0.19	0.44	0.12	0.18	0.17
GGF_05	0.54	0.20	0.87	0.06	0.16	0.15
GGF_14	0.06	0.18	0.25	0.08	0.24	0.15
GGF_12	0.16	0.21	0.74	0.10	0.16	0.18
GFA_01	0.12	0.22	0.46	0.17	0.16	0.17
GFA_02	0.12	0.37	0.53	0.07	0.22	0.22
GFA_08	0.06	0.10	0.18	0.08	0.11	0.32
GFA_16	0.10	0.13	0.41	0.08	0.22	0.18
AVERAGE	0.16	0.19	0.55	0.09	0.16	0.17

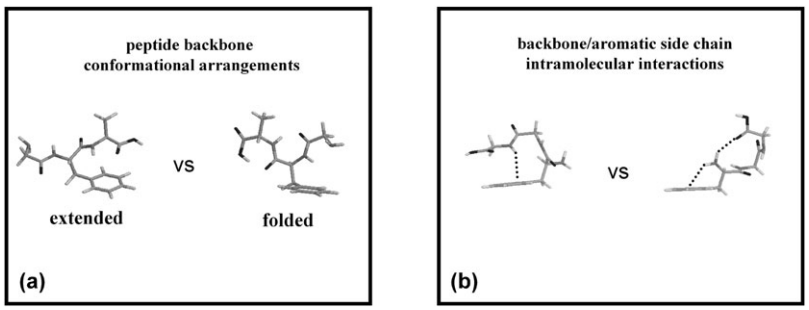


Fig. 1 Criteria of selection of structures used for the geometrical analysis.

conclusion that can be drawn from the analysis of Table 1 is the good performance shown by the tight-binding DFT-D method with an average RMSD value of 0.1 Å and each individual RMSD value smaller than 0.1 Å. This method, which is orders of magnitude faster than DFT, could be then applied either for (a) scanning the PES of a peptide and obtaining input geometries for a further *ab initio* treatment or (b) the geometry optimization of very large peptides for which the *ab initio* calculations are prohibited. The ff99 force field—independently of the set of charges used—performs also remarkably well with an average RMSD value similar to that of the TPSS-D or M06-2X functionals. Notice however, that the individual ff99 RMSDs values are systematically larger than 0.1 Å. This is also the case of the RMSDs values obtained with the M06-2X functional where the individual deviations oscillate between 0.1–0.2 Å with a maximum value of 0.46 Å for the WGG_03 structure. The M06-2X functional behaviour is more regular though than the one shown by the TPSS-D functional. For the latter, the discrepancies between the individual RMSDs values are much larger oscillating from the 0.06 Å (for GFA_08 and GGF_14) to 0.64 Å (for the FGG_215 structure). Notice however, that even for the structures with the largest RMSD values, the geometrical differences between the MP2 benchmark geometries and the TPSS-D or M06-2X ones are not that relevant, meaning that the overall geometry is essentially the same (see *e.g.* structure FGG_215 in Fig. 2(a) and structure WGG_03 in Fig. 2(b)). Therefore, the performance of both functionals is satisfactory.

One important remark should be done concerning the performance of the TPSS-D and M06-2X functionals compared to the MP2 method. This is the fact that the geometrical differences arise mainly from a slightly different orientation of the aromatic side chain with respect to the peptide backbone—where the London dispersion forces are mainly controlling the backbone-aromatic side chain intramolecular interactions—and not from the backbone conformational arrangements—where H-bonds are dominant (see Fig. 2(a) and (b)). Then, these geometrical differences should also be partially affected by the overestimation of the dispersion energy by the MP2 method.[47,48] Finally, the B3LYP functional gives by far the largest deviations—from all the methods considered—with respect to the MP2 geometries giving RMSDs values systematically larger than 0.2 Å (see Table 1 and Fig. 2(c) and (d) for an illustrative example). Obviously, this failure is due to the improper description of the dispersion energy provided by this functional.

Energetic analysis

The energetic analysis has been performed on the database built as specified in the Methods and computational details section. Thus, unlike in the geometrical analysis, here we have considered a total of 76 conformations. Then, for each level of theory of each individual peptide we have taken as a reference the most stable conformer and subsequently, the relative energies of the remaining conformers have been calculated. We will start discussing the data obtained from the statistical analysis based on the mean unsigned errors (MUE), standard deviations (σ) and maximum MUE (max) obtained comparing for each individual conformer of each individual peptide the CCSD(T)/CBS benchmark relative energies and the relative energies of the corresponding conformers calculated at the different levels of theory here considered.

Table 2 shows the following trend for the different methods assessed: ff99 > DFT (B3LYP, TPSS) > tight-binding DFT-D ~ DFT-D (PBE-D, BH&H, TPSS-D, M06-2X) > SCS-MP2 > MP2 > MP3, where the sign '>' means 'larger error than'.

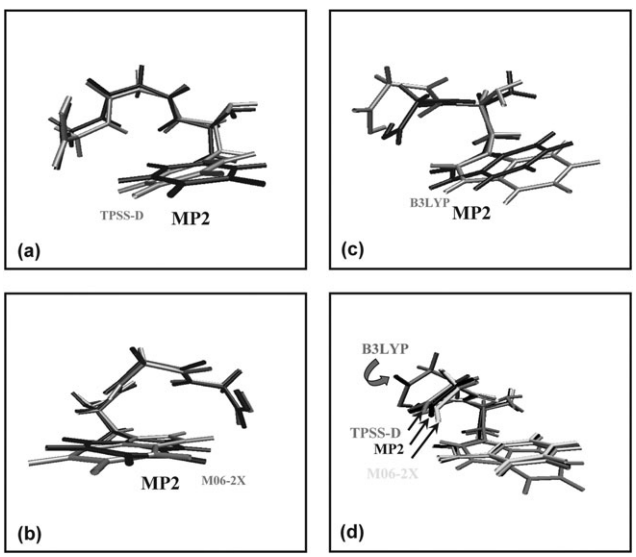


Fig. 2 (a) RI-MP2/cc-pVTZ (black) and TPSS-D/LP_1 (grey) geometries for the FGG_215 conformer. (b) RI-MP2/cc-pVTZ (black) and M06-2X/MIDI! (grey) geometries for the WGG_03 conformer. (c) RI-MP2/cc-pVTZ (black) and B3LYP/6-31G* (grey) geometries for the WG_03 conformer. (d) Comparison between RI-MP2/cc-pVTZ and, TPSS-D/LP_1, M06-2X/MIDI! and B3LYP/6-31G* geometries for the WG_01 conformer.

Notice that for the sake of simplicity, in the above classification, we have included the M06-2X and BH&H functionals in the DFT-D group, since they both belong to a last generation of functionals developed to improve the description of the dispersion energy given by other DFT functionals.

The performance of the ff99 force field is slightly better if the B3LYP/cc-pVTZ—instead of the HF/6-31G*—set of RESP charges is used but still, ff99 provides very inaccurate results. Notice, for instance, that for the WGG case, the maxima MUE (max) are 11.72 kcal/mol and 8.79 kcal/mol for the HF/6-31G* and B3LYP/cc-pVTZ set of RESP charges, respectively. Both, the B3LYP and TPSS functionals show a quite unsatisfactory behaviour specially when compared to the SCC-DF-TB-D method. Notice, that the tight-binding DFT-D method is not a pure DFT but a DFT-based method. The reason why the SCC-DF-TB-D method is superior to the B3LYP and TPSS functionals—which themselves are comparable—is undoubtedly its better description of the dispersion energy. This fact is also very well illustrated by the comparison of the performance shown by the TPSS-D and TPSS functionals. Augmenting the TPSS functional by an empirical term covering the London dispersion forces, improves significantly the quality of the method. Additionally, the performance of the tight-binding DFT-D method is very similar to that of the TPSS-D, PBE-D, M06-2X and BH&H functionals which in turn also show an analogous behaviour. However, the BH&H and PBE-D functionals typically show larger max and σ values than the tight-binding DFT-D and M06-2X methods. The TPSS-D functional performance depends on the system under study. In other words, it behaves similarly to the M06-

Table 2 Mean unsigned error (MUE), standard deviation (σ) and maximum MUE (max) obtained from the comparison between each of the CCSD(T)/CBS benchmark relative energies and the relative energies calculated at different levels of theory for each individual peptide. All energies are given in kcal/mol

	WFT			**DFT-D**		**SCC-DF-TB-D**
	MP2/CBS	SCS-MP2/CBS	MP3/CBS[a]	TPSS-D/LP$_1$[b]	PBE-D/LP$_1$[b]	
WG						
MUE	0.52	0.83	0.44	0.81	1.53	0.56
Σ	0.28	0.53	0.28	0.73	0.71	0.38
Max	0.92	1.55	0.89	1.80	2.45	1.44
WGG						
MUE	0.47	1.52	0.51	0.80	1.13	1.10
Σ	0.40	1.00	0.47	0.69	1.03	0.51
Max	1.15	3.14	1.51	2.78	3.69	2.06
FGG						
MUE	0.42	0.89	0.49	0.80	0.73	0.88
Σ	0.32	0.67	0.34	0.42	0.66	0.63
Max	1.13	2.35	1.11	1.54	2.31	2.16
GGF						
MUE	0.38	0.54	0.16	1.20	0.92	0.69
Σ	0.20	0.46	0.14	0.80	0.84	0.55
Max	0.78	1.21	0.40	2.75	2.70	1.83
GFA						
MUE	0.43	0.66	0.30	1.47	0.85	0.75
Σ	0.25	0.50	0.27	0.94	0.79	0.46
Max	0.98	1.69	0.81	3.02	2.58	1.42
	DFT	**ff99**				
	M06-2X/LP$_2$[c]	BH&H/LP$_3$[d]	B3LYP/LP$_1$[b]	TPSS/LP$_1$[b]	HF/6-31G*	B3LYP/cc-pVTZ
WG						
MUE	0.68	1.07	2.72	3.01	2.70	1.29
σ	0.50	0.71	1.35	1.71	1.98	0.87
Max	1.45	2.24	4.48	5.00	7.25	2.73
WGG						
MUE	0.62	1.39	3.15	3.06	3.41	2.28
σ	0.54	0.78	2.32	2.14	3.14	2.44
Max	1.71	2.87	6.67	7.08	11.72	8.79
FGG						
MUE	1.16	0.76	1.52	1.24	3.57	3.00
σ	0.65	0.50	1.27	1.27	1.93	1.59
Max	2.11	1.98	4.57	4.78	7.08	6.64
GGF						
MUE	0.58	1.13	1.24	1.56	4.99	3.86
σ	0.49	0.67	1.09	1.26	3.41	2.61
Max	1.67	2.13	4.42	4.59	10.93	8.17
GFA						
MUE	0.39	1.02	1.47	1.78	2.79	1.81
σ	0.21	0.84	0.91	1.21	2.20	1.32
Max	0.77	2.65	2.61	3.23	6.05	3.61

[a] $E_{CBS}^{MP3} = E_{CBS}^{MP2} + (E^{MP3} - E^{MP2})|_{6-31G^*}$. [b] LP$_1$ stands for 6-311++G(3df,3pd) Pople basis set. [c] LP$_2$ stands for 6-311+G(2df,2pd) Pople basis set. [d] LP$_3$ stands for 6-311++G(d,p) Pople basis set.

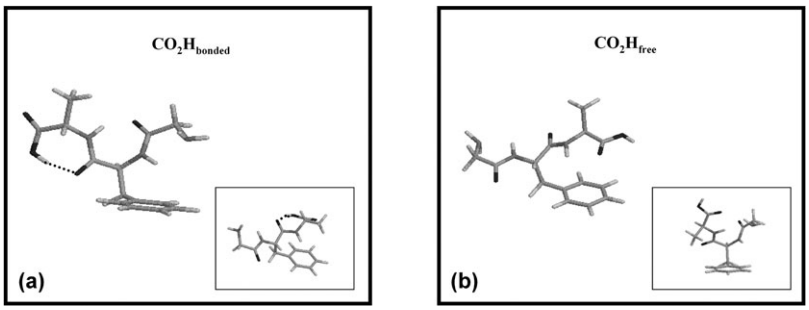


Fig. 3 Illustrative example of the two families of structures here considered. The small insets show that this classification is independent of the conformational arrangements of the peptide backbone.

2X functional and tight-binding DFT-D method for the WG and WGG systems whereas it behaves similar to the PBE-D and BH&H functionals for the remaining peptides. A possible explanation for this behaviour could be the different aromatic residue contained in the peptides here considered namely, a tryptophyl in case of WG and WGG and a phenylalanyl in case of FGG, GFA and GGF. Regarding the performance of the SCS-MP2, MP2 and MP3 methods, the following observations can be made. The SCS-MP2 method is comparable to the M06-2X functional and inferior to the MP2 and MP3 methods which themselves show—among all the levels of theory here considered—the smallest deviations with respect to the CCSD(T)/CBS benchmark data. Notice however that the SCS-MP2 relative energy values are typically smaller than the benchmark data whereas the MP2 values are larger since the first provides a better description of the dispersion energy, generally overestimated by the MP2 method (see Table S1 of the Supporting Information†). The MP3 method is superior to any of the MP2 methods although it overestimates the CCSD(T)/CBS relative energies (underestimates the stabilization energies) for the WG, WGG and GGF peptides whereas the opposite is true for the FGG and GFA systems. This suggests that the MP3 method performance is comparable to that of the CCSD method rather than to the CCSD(T) one.

Once we have discussed the global trends of performance of the different methods, we will analyze in more detail other relevant methodological aspects such as the localization of the global minimum, the ordering of structures within the considered set and the range of energy in which these structures are contained. This analysis is necessary because as already mentioned before, only a limited number of conformers—from the very extensive conformational landscape of a peptide—can be studied at the correlated quantum chemical level. Additionally, gas-phase experimental data on peptides reveal the coexistence of few conformers and thus, accurate theoretical calculations are needed in order to reproduce the experimental results. Since the selection of structures is based on their relative energies, it is essential then, that the method chosen for the theoretical study predicts correctly the global minimum, orders the structures properly and preferably, places them within similar intervals of energy. This last requirement is more desirable than essential because otherwise different energetic cutoffs should be considered for different methods in order to select the same number of conformers. For example, the 15 conformers of the WG dipeptide lie within an interval of energy of 2.45 kcal/mol at the CCSD(T)/CBS level of theory whereas the same set of structures lie within an interval of 7.34 kcal/mol at the TPSS/LP_1 level of theory (see Table S1 of the Supporting Information†). For the sake of simplicity of the analysis, we will take advantage of the fact that the different peptides conformations can be clustered into families. Then, we will classify the conformers depending on the existence or not of an intramolecular H-bond between the –OH of the carboxyl terminus group and the C═O group of the preceding residue (see Fig. 3). We will refer to these families as CO_2H_{bonded} (also indicated in inset(a) of Fig. 3 and in shady cells in Table S1†) and CO_2H_{free}, respectively. Notice that even when the London dispersion forces play a relevant role in the stability of all the conformers—independently to which family they belong—this effect is usually larger in those conformers of the CO_2H_{free} family, where the peptide backbone is typically lying over the aromatic side chain (see inset (b) in Fig. 3).

From the data collected in Table S1† it can be concluded that the ff99 force field—independently of the set of RESP charges used—fails in the prediction of the global minimum, in the ordering of the structures as well as in the calculation of the intervals of energy in which the conformers are placed. Likewise, the performance of the B3LYP and TPSS functionals is neither satisfactory. For example, in case of the WG dipeptide, conformers of the CO_2H_{bonded} (shady cells in Table S1†) and CO_2H_{free} families are more interspersed according to the DFT calculations than according to the CCSD(T)/CBS level of theory. Besides, the intervals of energy in which the structures are contained can differ significantly from the benchmark data, *e.g.*, for WGG tripeptide, these values are 10.1 kcal/mol and 9.26 kcal/mol for the B3LYP and TPSS functionals, respectively whereas at the CCSD(T)/CBS level of theory, the 15 structures are contained in 4.2 kcal/mol. The B3LYP and TPSS functionals also fail in the prediction of the global minimum. For instance, the global minimum according to the benchmark data for the FGG tripeptide is structure FGG_099. However, the same conformer has a relative energy of 4.57 kcal/mol and 4.78 kcal/mol at the B3LYP/LP_1 and TPSS/LP_1 levels of theory, respectively. Obviously then, the ordering of structures predicted by these functionals can be very different from the benchmark one. Notice for instance that for the GGF tripeptide, the least stable structure (GGF_01) at the CCSD(T)/CBS level of theory is placed in positions 11 and 9 according to the B3LYP and TPSS

calculations, respectively. Moreover, in the DFT scales, structure GGF_01 is approximately 3 kcal/mol more stable than structure GGF_05 whereas in the CCSD(T)/CBS scale, GGF_05 is 0.5 kcal/mol more stable than GGF_01.

The tight-binding DFT-D method predicts correctly the global minimum but typically structures are placed within an interval of energy approximately 1 kcal/mol larger than at the CCSD(T)/CBS level of theory. Besides, the ordering of structures can differ significantly from that of the benchmark data. For instance, for the WG dipeptide, the least stable conformer (WG_12) is placed in position number 7 according to the SCC-DF-TB-D scale. Similar failures are observed for the remaining peptides.

Regarding the M06-2X, BH&H, TPSS-D and PBE-D functionals the following observations can be made. First of all, as expected, they perform better than the B3LYP and TPSS functionals. Besides, the BH&H functional is comparable to the TPSS-D and PBE-D functionals. The BH&H, TPSS-D and PBE-D functionals show two tendencies of behaviour which are very well illustrated in the GFA and GGF cases (see Table S1†) namely, (a) generally, the conformers of the same family are grouped together and (b) the CO_2H_{free} family is usually predicted to be less stable than the CO_2H_{bonded} one. The reason for that is probably an underestimation of the dispersion energy included in these functionals. In case of the M06-2X functional, the conformers of the two families are more interspersed although there seems to be a subtle tendency to destabilize the CO_2H_{bonded} family (see for instance the GFA tripeptide). In other words, the M06-2X functional overstabilizes the conformers of the CO_2H_{free} family, suggesting that the description of the London dispersion forces provided by the M06-2X functional is still inaccurate. Regarding the intervals of energy in which the set of conformers is placed, the M06-2X functional behaves well compared with the reference data. Notice that the largest difference between the CCSD(T)/CBS and M06-2X/LP_2 maximum value of relative energy is 1.78 kcal/mol (for the FGG case), whereas the same value is 2.58 kcal/mol (GFA), 2.38 kcal/mol (GFA) and 2.55 kcal/mol (GGF) for the BH&H, TPSS-D and PBE-D functionals, respectively. Concerning the localization of the global minimum, M06-2X, BH&H, TPSS-D and PBE-D functionals predict typically the same global minima than CCSD(T)/CBS. However, there is a remarkable failure of the TPSS-D and PBE-D functionals for the FGG tripeptide, where the global minimum according to the benchmark calculations has a relative energy of 1.54 kcal/mol and 2.31 kcal/mol, respectively.

We will comment finally on the performance of the WFT methods here employed. MP2, SCS-MP2 and MP3 predict the same global minimum as calculated at the CCSD(T)/CBS level of theory. These methods also reproduce well the interval of energy in which the structures are contained (the difference in energy is systematically lower than 1 kcal/mol). The ordering of structures provided by the different methods is quite satisfactory as well although similarly to what has been discussed for the M06-2X functional, CO_2H_{bonded} and CO_2H_{free} conformers are more interspersed according to the SCS-MP2 method than according to the benchmark calculations. It is important to take into account though that for each peptide here considered few structures lie within an interval of energy smaller than 1 kcal/mol and then, it is difficult to establish a unique ranking. This is the case for instance of structure GFA_15. At the CCSD(T)/CBS and MP2/CBS levels of theory this structure is the global minimum whereas at the SCS-MP2/CBS and MP3/CBS levels of theory it is the third and second local minimum, respectively with a relative energy of 0.2 kcal/mol.

Additionally, a comment concerning the performance of the MP2, SCS-MP2 and MP3 methods for molecular clusters should be made.[49] The SCS-MP2 eliminates the overestimation of the dispersion energy and consequently, the method provides better results for stacked structures. Calculations carried out so far show that it is correct. But the SCS-MP2 method reduces also the stabilization energy for H-bonded structures (here it is not required) and for these structures, the SCS-MP2 gives too small stabilization energies. Likewise, the MP3 shows similar behaviour to the SCS-MP2 for stacked structures but its performance for H-bonded structures is better. The case of peptides is somehow different since the classical classification of molecular complexes as stacked and H-bonded structures is not applicable here because peptides are a combination of both ('H-bonded' in the peptide backbone and 'stacked' regarding the backbone/aromatic side chain interaction). Let us finally add that MP3 calculations require considerably more CPU time than either MP2 or SCS-MP2 method.

Conclusions

The main conclusions of this work can be summarized as follows:

(i) Neither geometries nor relative electronic energies obtained using AMBER (ff99 force field) are reliable. Thus, we do not recommend the use of this force field for the study of isolated small peptides containing aromatic side chains.

(ii) Tight-binding DFT-D (SCC-DF-TB-D) provides reasonably reliable geometries but the ordering of structures—according to their relative energies—differs from the benchmark one. Thus, SCC-DF-TB-D can be used, and so is recommended, for exploring the PES of an isolated small peptide containing aromatic side chains and localizing all the existing minima. However, the relative energies of these minima should be further recalculated with a more accurate methodology.

(iii) B3LYP and TPSS functionals are not recommended for the study of isolated small peptides containing aromatic side chains. Due to the missing London dispersion forces both, geometries and relative energies can be different from the benchmark data.

(iv) M06-2X, BH&H, TPSS-D and PBE-D show a reasonably good performance when compared to the CCSD(T)/CBS benchmark data. M06-2X and TPSS-D behave similar to some extent. When compared to the benchmark data, both give reasonably good geometries but major differences are found regarding the ordering of structures. Keeping in mind that the CPU time requirements needed by these functionals are still lower than the MP or CC methods, and mainly, when thinking about the theoretical study of larger systems than tripeptides, we recommend the use of these functionals—or similar—for

the reoptimization of the most stable minima localized in the SCC-DF-TB-D PES. However, the further selection of the most stable structures in the PES should be based on single-point energy calculations performed on the M06-2X and/or TPSS-D geometries with either the MP or the CC methods as explained in the following paragraph. Additionally, the CPU time requirements of the RI-DFT-D(TPSS-D) calculations are considerably lower (by one order of magnitude) than that of the M06-2X calculations. Consequently, the RI-DFT-D procedure unlike the M06-2X functional can be also used for carrying out on-the-fly MD simulations.

(v) MP3/CBS, MP2/CBS and SCS-MP2/CBS single-point energy calculations provide orderings of structures reliable enough when compared to the CCSD(T) CBS one. Consequently, it is possible to make the final selection of conformers based on these ranking of energies. Besides, it is important to stress that the extrapolation of the basis set to the complete basis set limit is strictly necessary as it guarantees that the relative energies are not affected by the intramolecular basis set superposition error.

(vi) From all the conclusions commented above our recommendation for the theoretical study of isolated small peptides containing an aromatic side chain is the following: the PES of each individual peptide should be carefully explored by means of the tight-binding DFT-D methodology. Then, few of the most stable conformers should be selected and further reoptimized using either the RI-MP2 method or, in case of larger systems than tripeptides, the M06-2X or the TPSS-D functionals. Finally, single-point energy calculations should be carried out on the latter geometries either at the MP3/CBS, MP2/CBS or the SCS-MP2/CBS levels of theory (in this order of preference). The order of structures obtained at these levels of theory should be reliable enough as to make the final selection of structures.

Acknowledgements

This work was supported by Grant No. A400550510 and LC512 from the Grant Agency of the Academy of Sciences of the Czech Republic and MSMT of the Czech Republic; it was also part of the research project No. Z40550506. A portion of the research described in this paper was performed in the Environmental Molecular Sciences Laboratory, a national scientific user facility sponsored by the Department of Energy's Office of Biological and Environmental Research and located at Pacific Northwest National Laboratory.

References

1 F. Ryjacek, O. Engkvist, J. Vacek, M. Kratochvil and P. Hobza, *J. Phys. Chem. A*, 2001, **105**(7), 1197–1202.
2 D. Reha, H. Valdes, J. Vondrasek, P. Hobza, A. bu-Riziq, B. Crews and M. S. de Vries, *Chem.–Eur. J.*, 2005, **11**(23), 6803–6817.
3 H. Valdes, V. Spiwok, J. Rezac, D. Reha, A. G. Abo-Riziq, M. S. de Vries and P. Hobza, *Chem.–Eur. J.*, 2008, in press.
4 M. Elstner, P. Hobza, T. Frauenheim, S. Suhai and E. Kaxiras, *J. Chem. Phys.*, 2001, **114**(12), 5149–5155.
5 M. Iannuzzi, A. Laio and M. Parrinello, *Phys. Rev. Lett.*, 2003, **90**(23).
6 A. Laio and M. Parrinello, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**(20), 12562–12566.
7 J. M. Bakker, C. Plutzer, I. Hunig, T. Haber, I. Compagnon, G. von Helden, G. Meijer and K. Kleinermanns, *ChemPhysChem*, 2005, **6**(1), 120–128.
8 W. Chin, M. Mons, J. P. Dognon, F. Piuzzi, B. Tardivel and I. Dimicoli, *Phys. Chem. Chem. Phys.*, 2004, **6**(10), 2700–2709.
9 H. Fricke, G. fer, T. Schrader and M. Gerhards, *Phys. Chem. Chem. Phys.*, 2007, **9**(32), 4592–4597.
10 J. M. Wang, P. Cieplak and P. A. Kollman, *J. Comput. Chem.*, 2000, **21**(12), 1049–1074.
11 P. Jurecka, J. Cerny, P. Hobza and D. R. Salahub, *J. Comput. Chem.*, 2007, **28**(2), 555–569.
12 J. P. Perdew and Y. Wang, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1992, **45**(23), 13244–13249.
13 J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**(18), 3865–3868.
14 J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1997, **78**(7), 1396.
15 Y. Zhao and D. G. Truhlar, *Theor. Chem. Acc.*, 2007, submitted.
16 C. T. Lee, W. T. Yang and R. G. Parr, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1988, **37**(2), 785–789.
17 J. M. Tao, J. P. Perdew, V. N. Staroverov and G. E. Scuseria, *Phys. Rev. Lett.*, 2003, **91**(14).
18 A. D. Becke, *Phys. Rev. A: At., Mol., Opt. Phys.*, 1988, **38**(6), 3098–3100.
19 Chr. Møller and S. Plesset, *Phys. Rev.*, 2007, **46**, 618–622.
20 S. Grimme, *J. Chem. Phys.*, 2003, **118**(20), 9095–9102.
21 W. J. Hehre, L. Radom, P. v. R. Schleyer and J. A. Pople, in *Ab Initio Molecular Orbital Theory*, Wiley-Interscience, New York, 1986.
22 R. E. Easton, D. J. Giesen, A. Welch, C. J. Cramer and D. G. Truhlar, *Theor. Chim. Acta*, 1996, **93**(5), 281–301.
23 R. A. Kendall, T. H. Dunning and R. J. Harrison, *J. Chem. Phys.*, 1992, **96**(9), 6796–6806.
24 L. F. Holroyd and T. van Mourik, *Chem. Phys. Lett.*, 2007, **442**(1–3), 42–46.
25 H. Valdes, V. Klusak, M. Pitonak, O. Exner, I. Stary, P. Hobza and L. Rulisek, *J. Comput. Chem.*, 2007, DOI 10.1002/jcc.
26 T. van Mourik, P. G. Karamertzanis and S. L. Price, *J. Phys. Chem. A*, 2006, **110**(1), 8–12.
27 H. Valdes, D. Reha and P. Hobza, *J. Phys. Chem. B*, 2006, **110**(12), 6385–6396.
28 M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai and G. Seifert, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1998, **58**(11), 7260–7268.
29 K. Eichkorn, O. Treutler, H. Ohm, M. Haser and R. Ahlrichs, *Chem. Phys. Lett.*, 1995, **240**(4), 283–289.
30 K. Eichkorn, F. Weigend, O. Treutler and R. Ahlrichs, *Theor. Chem. Acc.*, 1997, **97**(1–4), 119–124.
31 J. Cizek, *J. Chem. Phys.*, 1966, **45**(11), 4256.
32 J. Cizek and J. Paldus, *Int. J. Quantum Chem.*, 1972, **6**, 435–438.
33 A. Halkier, T. Helgaker, P. Jorgensen, W. Klopper, H. Koch, J. Olsen and A. K. Wilson, *Chem. Phys. Lett.*, 1998, **286**(3–4), 243–252.
34 P. Jurecka and P. Hobza, *Chem. Phys. Lett.*, 2002, **365**(1–2), 89–94.
35 M. Kabelac, H. Valdes, E. C. Sherer, C. J. Cramer and P. Hobza, *Phys. Chem. Chem. Phys.*, 2007, **9**(36), 5000–5008.
36 J. Cerny, P. Jurecka, P. Hobza and H. Valdes, *J. Phys. Chem. A*, 2007, **111**(6), 1146–1154.
37 C. I. Bayly, P. Cieplak, W. D. Cornell and P. A. Kollman, *J. Phys. Chem.*, 1993, **97**(40), 10269–10280.
38 B. Aradi, B. Hourahine and T. Frauenheim, *J. Phys. Chem. A*, 2007, **111**(26), 5678–5684.
39 R. D. Amos, A. Bernhardsson, A. Berning, P. Celani, D. L. Cooper, M. J. O. Deegan, A. J. Dobbyn, F. Eckert, C. Hampel, G. Hetzer, P. J. Knowles, T. Korona, R. Lindh, A. W. Lloyd, S. J. McNicholas, F. R. Manby, W. Meyer, M. E. Mura, A. Nicklass, P. Palmieri, R. Pitzer, G. Rauhut, M. Schütz, U. Schumann, H. Stoll, A. J. Stone, R. Tarroni, T. Thorsteinsson and H.-J. Werner, *MOLPRO, a package of* ab initio *programs designed by H.-J. Werner and P. J. Knowles, Version 2002.1*, 2002.
40 G. Karlstrom, R. Lindh, P. A. Malmqvist, B. O. Roos, U. Ryde, V. Veryazov, P. O. Widmark, M. Cossi, B. Schimmelpfennig, P. Neograd and L. Seijo, *Comput. Mater. Sci.*, 2003, **28**(2), 222–239.
41 P. Neograd, F. Aquilante, J. Noga, M. Pitonak and M. Urban, unpublished work.

42 R. Ahlrichs, M. Bar, M. Haser, H. Horn and C. Kolmel, *Chem. Phys. Lett*., 1989, **162**(3), 165–169.
43 Y. Shao, L. F. Molnar, Y. Jung, J. Kussmann, C. Ochsenfeld, S. T. Brown, A. T. B. Gilbert, L. V. Slipchenko, S. V. Levchenko, D. P. O'Neill, R. A. DiStasio, R. C. Lochan, T. Wang, G. J. O. Beran, N. A. Besley, J. M. Herbert, C. Y. Lin, T. Van Voorhis, S. H. Chien, A. Sodt, R. P. Steele, V. A. Rassolov, P. E. Maslen, P. P. Korambath, R. D. Adamson, B. Austin, J. Baker, E. F. C. Byrd, H. Dachsel, R. J. Doerksen, A. Dreuw, B. D. Dunietz, A. D. Dutoi, T. R. Furlani, S. R. Gwaltney, A. Heyden, S. Hirata, C. P. Hsu, G. Kedziora, R. Z. Khalliulin, P. Klunzinger, A. M. Lee, M. S. Lee, W. Liang, I. Lotan, N. Nair, B. Peters, E. I. Proynov, P. A. Pieniazek, Y. M. Rhee, J. Ritchie, E. Rosta, C. D. Sherrill, A. C. Simmonett, J. E. Subotnik, H. L. Woodcock, W. Zhang, A. T. Bell, A. K. Chakraborty, D. M. Chipman, F. J. Keil, A. Warshel, W. J. Hehre, H. F. Schaefer, J. Kong, A. I. Krylov, P. M. W. Gill and M. Head-Gordon, *Phys. Chem. Chem. Phys*., 2006, **8**(27), 3172–3191.
44 D. A. Case, T. A. Darden, T. E. Cheatham III, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, K. M. Merz, D. A. Pearlman, M. Crowley, R. C. Walker, W. Zhang, B. Wang, S. Hayik, A. Roitberg, G. Seabra, K. F. Wong, F. Paesani, X. Wu, S. Brozell, V. Tsui, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, P. Beroza, D. H. Mathews, C. Schafmeister, W. S. Ross and P. A. Kollman, *AMBER 9*, University of California, San Francisco, 2006.
45 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. G. Johnson, W. Chen, M. W. Wong, C. Gonzalez and J. A. Pople, *GAUSSIAN 03 (Revision C.02)*, Gaussian, Inc., Wallingford, CT, 2004.
46 P. Jurecka, J. Sponer, J. Cerny and P. Hobza, *Phys. Chem. Chem. Phys*., 2006, **8**(17), 1985–1993.
47 G. J. O. Beran, M. Head-Gordon and S. R. Gwaltney, *J. Chem. Phys*., 2006, **124**(11).
48 P. Hobza, H. L. Selzle and E. W. Schlag, *J. Phys. Chem*., 1996, **100**(48), 18790–18794.
49 M. Pitonak, K. E. Riley and P. Hobza, unpublished work.