Journal of Molecular Graphics and Modelling 29 (2010) 178–187

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

# Evaluation of dispersion-corrected density functional theory (B3LYP-DCP) for compounds of biochemical interest

Sten O. Nilsson Lill*

*Department of Chemistry, University of Gothenburg, SE-412 96 Göteborg, Sweden*

ARTICLE INFO

*Article history:*
Received 21 April 2010
Received in revised form 1 June 2010
Accepted 2 June 2010
Available online 15 June 2010

*Keywords:*
Density functional theory
Dispersion correction
Aromatic interactions
Peptide
Conformational analysis

ABSTRACT

An evaluation of a dispersion-corrected density functional theory method (B3LYP-DCP) [I.D. Mackie, G.A. DiLabio, Interactions in large, polyaromatic hydrocarbon dimers: application of density functional theory with dispersion corrections, J. Phys. Chem. A 112 (2008) 10968–10976] for three systems of biochemical interest is presented. Firstly, structures and energies of isomers of the tripeptide Phe-Gly-Phe have been compared with CCSD(T)/CBS//RI-MP2/cc-pVTZ literature values. In the system aromatic interactions compete with XH–π (X = C, N) interactions and hydrogen bonds which makes it a reliable model for proteins. The resulting mean absolute deviation between B3LYP-DCP and CCSD(T)/CBS relative energies is found to be 0.50 kcal mol$^{-1}$. Secondly, a phenylalanine derivative featuring a CH–π interaction has been investigated. A comparison between the optimized geometry and X-ray crystal data shows that B3LYP-DCP accurately predicts the interaction between the two aromatic rings. Thirdly, the dipeptide Ac-Phe-Phe-$NH_2$ which contains an edge-to-face interaction between two aromatic rings has been studied. The study demonstrates the general applicability of the B3LYP-DCP method on systems which features interactions typically present in biochemical compounds.



## 1. Introduction

In the last decades it has become more and more evident that weak interactions play significant roles in biological recognition. Examples of this are interactions between aromatic residues in enzymes in general, interactions between drugs and aromatic residues in the active site of an enzyme, interactions between carbohydrates and the protein matrix, XH–π (X = C, N, O, S) interactions, and hydrogen bonds [1–11]. A very illustrative example of this is the blend of interactions found between the anti-Alzheimer drug Aricept (2-((1-benzylpiperidin-4-yl)methyl)-5,6-dimethoxy-2,3-dihydroinden-1-one) and the enzyme acetylcholinesterase (Fig. 1) [2].

One of the major challenges in computational chemistry is to be able to handle weak interactions such as hydrogen bonds and dispersive interactions in an accurate way. Although a high-level wave function method such as CCSD(T) is capable of accurately describing these types of interactions it is generally not applicable to systems of large size due to its inherent demand of computational resources. Therefore, for the less time-consuming density functional theory (DFT) methods, considerable effort has been put into the development of correction terms to treat dispersive interactions between non-covalently bonded molecules. For example, the group of DiLabio has developed dispersion-correcting atom-centred potentials (DCPs) for use with a variety of common DFT functionals and basis sets [12,13]. The DCP is added in the form of two Gaussian functions, one long-range attractive and a second short-ranged repulsive. For larger systems this was found necessary to avoid over-binding [12,13]. The coefficients for the two functions were optimized to reproduce the dispersion binding in a training set consisting of non-covalently bound dimers of hydrocarbons and of heteroatom containing molecules. It was found that binding energies could be predicted within 15%, and monomer separations within 0.1 Å, of high-level *ab initio* data [12,13]. DCPs have successfully been applied to study dispersive binding in a diverse set of systems, including cyclic aromatic hydrocarbons [12–14], benzothiophenes [15], small hydrocarbons and pentacene on hydrogen-terminated silicon surfaces [16,17], and recently also for hydrocarbon isomerisation reactions and olefin monomer insertion reactions [18]. The DCP approach is attractive in three ways: it gives the possibility of switching the correction term "on or off" and therefore directly gives the answer to how important dispersive interactions are for the system under investigation. DCPs are also applicable in a number of general computational chemistry programs, and one does not need specialised functionals or large basis sets [12,13]. An alternative way of handling dispersive interactions has been described by Grimme and co-workers. They have developed DFT-D in which a damped empirical atom pair specific

* Tel.: +46 31 7869103; fax: +46 31 7723840.
*E-mail address:* stenil@chem.gu.se.

1093-3263/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2010.06.001

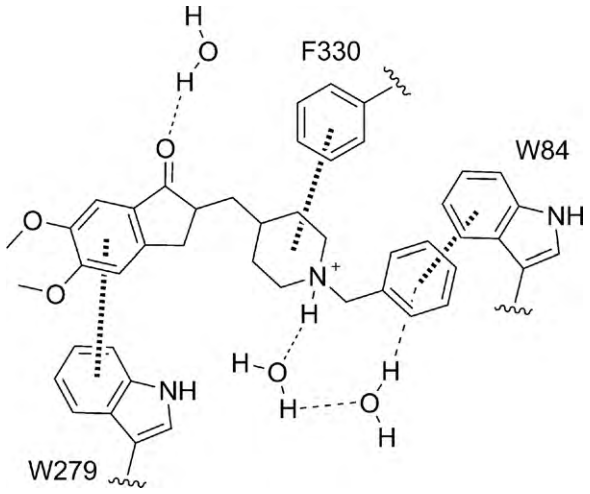


**Fig. 1.** Interactions (dashed lines) between Aricept and the active site of acetylcholinesterase.

potential of $C_6/R^6$ type is added on top of existing DFT functionals [19–21]. As a comparison, with DFT-D the dispersion attraction stems from attraction between nuclei, while in the DCP-approach the attraction comes from the added potential that acts between electrons and nuclei. Using the successful DFT-D approach one can also switch the correction term "on or off" and directly estimate its influence on energies and geometries. Besides these techniques, new DFT functionals in which weak interactions are well described compared to how standard DFT functionals perform have recently been released, for example M06 and M06-2X [22]. In addition, truly non-local DFT-methods, such as the vdW-DF approach based on the Andersson–Langreth–Lundqvist (ALL)-functional has shown great promise [23,24].

B3LYP, which is a standard DFT-approach, in general shows great performance for thermochemistry [25,26], but due to its repulsive long-range behavior [12] B3LYP is not recommended for weakly interacting systems [26]. Recently [14], the dispersion-corrected B3LYP/6-31+G(d,p)-DCP method was evaluated against the S22 benchmark set [27] which contains a range of complexes interacting via dispersive interactions, hydrogen bonds, or combinations of these. With a mean deviation of only 0.8 kcal $mol^{-1}$ to benchmark wave function calculations for dimerization energies, it was shown that B3LYP/6-31+G(d,p)-DCP is a useful method for weakly interacting systems [14]. The 6-31+G(d,p) basis set used was the one recommended by the DCP developers to give a good balance between computing time and good quality results [13], and if a method with such a fairly small basis set gives reliable results it would make it very useful for studies of systems of larger size. However, it is of interest to continue to evaluate B3LYP/6-31+G(d,p)-DCP also for systems of biochemical interest to demonstrate the general applicability of the method. In this study, the following systems have been investigated which separates the paper in three sections:

(a) The tripeptide Phe-Gly-Phe for which high level *ab initio* energies and structures are available [28]. This allows for an adequate evaluation of different isomers calculated using B3LYP-DCP. In the isomers a competition between aromatic ring-ring interactions, XH–π (X = C, N) interactions, and hydrogen bonds occurs.
(b) A phenylalanine derivative with available X-ray structural data demonstrating an intramolecular CH–π interaction [29]. This allows for direct structural validation of B3LYP-DCP on this type of weak interaction. In addition, $^1$H NMR data, sensitive to structural changes and therefore useful chemical probes, are evaluated.
(c) The dipeptide Ac-Phe-Phe-$NH_2$, which has been characterized in the gas-phase by IR-spectroscopy and computational studies [30]. This system shows an edge-to-face interaction between the two aromatic rings. A comparison of calculated and experimental vibrational frequencies, also sensitive to different interactions, is made.

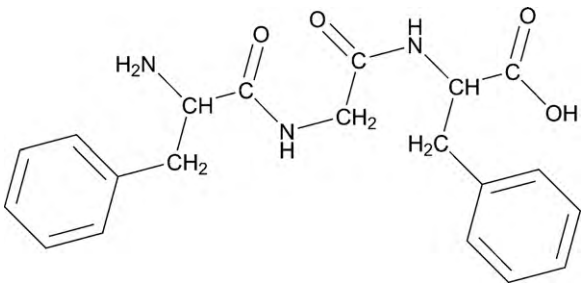

**Fig. 2.** The tripeptide system Phe-Gly-Phe.

## 2. Computational methods

All compounds were optimized using B3LYP[31–33]/6-31+G(d,p)[34–36]-DCP [13] in Gaussian03 [37]. Examples of input for performing this type of calculation using Gaussian can be found in the Supporting Information. Since the dispersion correcting potential as explained in the original references explicitly was parameterized to cover the basis set superposition error, no such correction was employed here [12,13]. All geometries were characterized as minima or saddle points on the potential energy surface by using the sign of the eigenvalues of the force constant matrix obtained from a frequency calculation. Single point calculations for the tripeptide system were also performed using B3LYP/6-31+G(d,p)-DCP//RI-MP2/cc-pVTZ for a comparison. The MP2 geometries were taken from Ref. [28]. Chemical shieldings, *i.e.* the second derivative of the energy with respect to an external magnetic field and the nuclear magnetic moment, were calculated with the standard Gauge Independent Atomic Orbital (GIAO) [38] method using B3LYP/6-31+G(d,p), both with and without the DCP included, or with the larger triple zeta basis set 6-311+G(d,p)-DCP, relative to TMS optimized in $T_d$-symmetry. Figures and plots of calculated structures were generated using Chemcraft [39].

## 3. Results and discussion

### 3.1. The tripeptide system Phe-Gly-Phe

In a recent study by Valdes et al. the tripeptide Phe-Gly-Phe was investigated using benchmark CCSD(T)/CBS//RI-MP2/cc-pVTZ calculations (Fig. 2) [28].

The tripeptide system includes aromatic side-chains which are capable of forming stacked or T-shaped complexes due to stabilizing interactions between aromatic rings. In addition it may form CH–π and NH–π interactions and hydrogen bonds of NH···O=C, OH···O=C, or NH···$NH_2$ type, so it appears to be a realistic model system for competitive interactions typically observed in proteins. In this study, we take advantage of the recent data from Valdes et al.'s study [28] on different isomers of the tripeptide Phe-Gly-Phe and compare with results generated using B3LYP-DCP.

In the study by Valdes et al., different computational approaches were evaluated against CCSD(T)/CBS//RI-MP2/cc-pVTZ calculations for fifteen different conformers (Fig. 3) [28]. These were ranked in terms of relative potential energies and RMS structural devi-

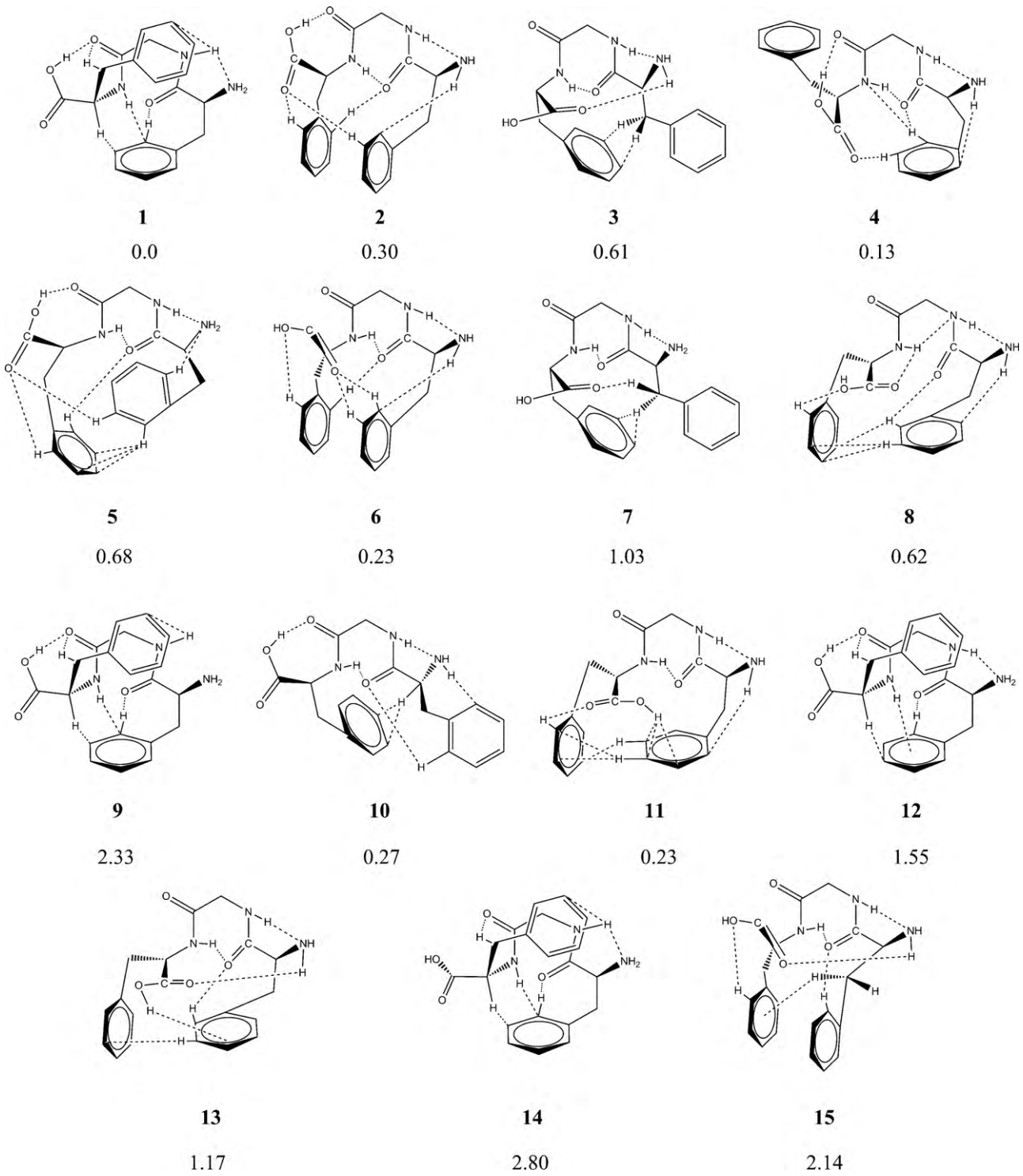


**Fig. 3.** Schematic illustration of isomers. Relative energies are given in kcal mol$^{-1}$.

ations compared to CCSD(T)/CBS//RI-MP2/cc-pVTZ results. In this study, a similar procedure and conformer numbering is adopted for the most easy and transparent comparison between the studies. In the relative energy evaluation two different statistical evaluation methods were tested (Evaluations A and B in Table 1).

Among the 15 different isomers, there are three (**2**, **6**, **15**) which feature a slipped stacked arrangement of the aromatic phenyl rings. In addition, there are four isomers (**5**, **8**, **11**, **13**) where the rings are oriented in a T-shaped manner, while the remaining eight isomers have only classical hydrogen bonds or covalent bonds. From the plot of relative energies ($\Delta E$) for the different isomers (Fig. 4), it is gratifyingly observed that both B3LYP-DCP and CCSD(T) identify isomer **1** as the most stable.

In addition, the energy variation among the 15 isomers is well described by B3LYP-DCP (2.80 kcal mol$^{-1}$) compared to the variation from the CCSD(T) calculations (3.19 kcal mol$^{-1}$). The four least stable isomers (**12**–**15**) predicted by CCSD(T) are also among the five least stable isomers employing B3LYP-DCP (*cf.* Fig. 3 for relative energies). As can be seen, the relative energies between isomer **1** and the other are generally smaller for B3LYP-DCP than

**Table 1**
Mean absolute deviation (MAD in kcal mol$^{-1}$) of differences in relative energies between different methods and CCSD(T)/CBS calculations.

| Statistical evaluation | Method | Overall | S-complexes | T-complexes | (S+T)-complexes | Other |
|---|---|---|---|---|---|---|
| A[a] | B3LYP-DCP//B3LYP-DCP | 0.88 | 1.01 | 1.28 | 1.17 | 0.62 |
| B[b] | | 0.50 | 0.26 | 0.49 | 0.39 | 0.59 |
| A | B3LYP-DCP//RI-MP2/cc-PVTZ | 0.88 | 1.15 | 1.37 | 1.28 | 0.53 |
| B | | 0.61 | 0.38 | 0.59 | 0.50 | 0.71 |
| A | B3LYP/6-311++G(3df,3pd)//RI-MP2/cc-pVTZ | 1.82[c] | | | | |
| B | | 1.94 | | | | |

[a] Relative energy evaluation compared to only isomer **1**.
[b] Relative energy evaluation compared to the average of all 15 isomers.
[c] Calculated using energies extracted from the Supporting Information of Ref. [28].

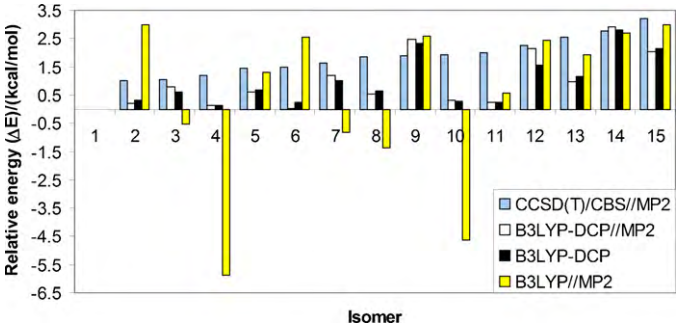


**Fig. 4.** Relative energies for different isomers (**1**–**15**) of the tripeptide Phe-Gly-Phe. Isomer **1** is set as the reference point.

for CCSD(T). Only isomers **9** and **14** are described as somewhat too high in energy by the DFT-approach, but with only small deviations from the benchmark values. Overall, it is found that B3LYP-DCP is capable of describing relative energies among different isomers of polypeptides with competitive hydrogen bonds and π-interactions involved. This is emphasized by examining also the performance of uncorrected B3LYP (Fig. 4) which does not even reproduce isomer **1** as the most stable isomer. Also, the energy variation among the isomers is much larger (8.88 kcal mol$^{-1}$) than that observed for B3LYP-DCP, even though the B3LYP energies were calculated on RI-MP2/cc-pVTZ optimized geometries and with a large flexible basis set (6-311++G(3df,3pd)) [28].

Fig. 5 illustrates, for each isomer, the difference in relative energy ($\Delta\Delta E$) between the different computational approaches used and Valdes' CCSD(T) results ($\Delta\Delta E = \Delta E$(CCSD(T)) − $\Delta E$(B3LYP-DCP)).

A positive $\Delta\Delta E$-value for an isomer means that the DFT calculations result in a smaller $\Delta E$ than the corresponding $\Delta E$ from the CCSD(T)-calculations. Thus, DFT is over-stabilizing such an isomer compared to isomer **1**, in relation to the CCSD(T) reference data. In Fig. 5 is also inserted the line for the mean absolute deviation (MAD = 0.88 kcal mol$^{-1}$) in $\Delta\Delta E$ using B3LYP-DCP for all the isomers for an easier identification of which isomers improve or

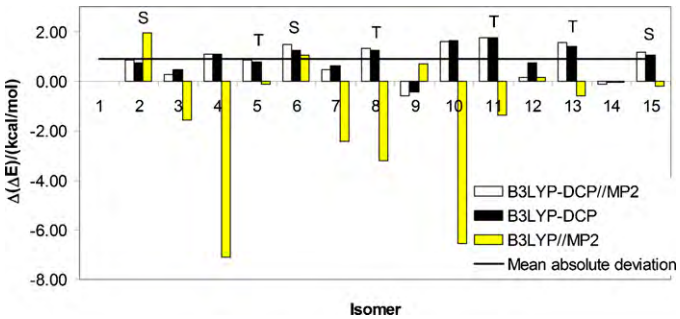


**Fig. 5.** Difference in relative energies ($\Delta(\Delta E)$) between CCSD(T) and B3LYP-DCP or B3LYP calculations. S and T denote stacked and T-shaped complexes, respectively.

deteriorate the deviation, respectively (Evaluation A). Interestingly, three out of the four T-shaped complexes and two out of the three stacked complexes show a larger deviation than the overall MAD. The largest deviation among all isomers is found for the T-shaped isomer **11** where a $\Delta\Delta E$ value of 1.75 kcal mol$^{-1}$ is found. Thus, at first it appears that the B3LYP-DCP method in general overestimates the stabilizing interaction for the stacked and T-shaped complexes in comparison to the CCSD(T) relative energies. This was a bit unexpected since both the S22 benchmark comparison [14] and the results from Mackie and DiLabio's study [13] rather showed that the interaction energies for T-shaped complexes were underestimated. Also, in the study of the toluene dimer, the T-shaped complex reorganized to a slipped stacked isomer using B3LYP-DCP [14]. An alternative explanation could be that the reference point, isomer **1**, is described as being too unstable with B3LYP-DCP compared to CCSD(T)/CBS calculations. The mean deviation in $\Delta\Delta E$ for the stacked and T-shaped complexes is found to be 1.17 kcal mol$^{-1}$, which is 0.29 kcal mol$^{-1}$ higher than the overall MAD (Table 1). This results in a deviation of 0.62 kcal mol$^{-1}$ for the eight remaining isomers (Table 1).

To avoid the dependence on a single isomer in the comparison and evaluation of the relative energies a modified strategy was introduced in Valdes' paper and also tested here (Evaluation B in Table 1). Instead of comparing with only the most stable isomer, the average of all 15 isomers is used. If following this approach, the MAD between B3LYP-DCP and CCSD(T)/CBS calculated relative energies is calculated to be 0.50 kcal mol$^{-1}$, which is to be compared with the MAD of 0.54 and 0.51 kcal mol$^{-1}$ for M06-2X/MIDI! and SCC-DF-TB-D, respectively, reported in Valdes et al's paper [28]. Thus, these three methods all equally well describe the relative energies of the different isomers of the Phe-Gly-Phe tripeptide. With this analysis, the S- and T-complexes have a lower MAD than the other complexes, and thus are generally better described than other isomers in opposite to the previous analysis. Thus, for the analysis and comparison of relative energies one needs to be cautious. As a comparison, the results from uncorrected B3LYP calculations are also included in Table 2. Without the dispersion correcting potential, the deviation is demonstrated to be significantly larger (1.94 kcal mol$^{-1}$).

The deviation for the stacked and T-shaped complexes was also checked not to be an effect of the B3LYP-DCP optimized geometries. A comparison of relative energies following Evaluation A, but using B3LYP/6-31+G(d,p)-DCP//RI-MP2/cc-PVTZ energies, thus using MP2 geometries, resulted also in an overall MAD of 0.88 kcal mol$^{-1}$. The deviation for all stacked and T-shaped complexes is slightly higher (1.28 kcal mol$^{-1}$ vs. 1.17 kcal mol$^{-1}$, Table 1) with the T-shaped complexes contributing by 1.37 kcal mol$^{-1}$ and stacked complexes by 1.15 kcal mol$^{-1}$. The deviation for the remaining eight complexes is lowered to 0.53 kcal mol$^{-1}$. Similarly, applying Evaluation B in the analysis gives an overall MAD of 0.61 kcal mol$^{-1}$ with a generally lower deviation for S- and T-complexes. Thus, B3LYP-DCP//B3LYP-DCP energies result in a smaller deviation for S- and T-complexes compared to B3LYP-DCP//RI-MP2/cc-PVTZ ener-

**Table 2**
Root-mean squared deviations (RMSD, in Å) between geometries of the tripeptide Phe-Gly-Phe optimized using B3LYP/6-31+G(d,p)-DCP or RI-MP2/cc-PVTZ.

| Isomer | Complex type | RMSD |
|---|---|---|
| 1 | | 0.09 |
| 2 | S | 0.07 |
| 3 | | 0.28 |
| 4 | | 0.14 |
| 5 | T | 0.08 |
| 6 | S | 0.08 |
| 7 | | 0.21 |
| 8 | T | 0.11 |
| 9 | | 0.07 |
| 10 | | 0.20 |
| 11 | T | 0.28 |
| 12 | | 0.33 |
| 13 | T | 0.06 |
| 14 | | 0.11 |
| 15 | S | 0.14 |
| Average: 1–15 | | 0.15 |

gies, while the opposite is true for the other eight complexes. This indicates that the deviation between B3LYP-DCP and CCSD(T)/CBS results has its origin from the energy evaluation and not from the geometry used. The MAD between $\Delta E$:*s* calculated using B3LYP-DCP//B3LYP-DCP or B3LYP-DCP//RI-MP2/cc-PVTZ geometries is as small as 0.14 kcal mol$^{-1}$ with the largest deviation of 0.59 kcal mol$^{-1}$ for isomer **12**. The small differences in relative energy when varying the geometry used suggests that the geometries are very alike. To confirm this, root-mean squared deviations (RMSD) were calculated between B3LYP-DCP and RI-MP2/cc-PVTZ geometries for each isomer. These values are listed in Table 2. The average of the RMSD-value was found to be 0.15 Å. This can be compared with the two best methods in Valdes' study: M06-2X/MIDI! and SCC-DF-TB-D which resulted in average RMSD-values of 0.10 and 0.08 Å, respectively. Thus, although B3LYP-DCP is slightly less agreeing with RI-MP2/cc-PVTZ geometries than the other two methods it will be a valuable alternative, especially if programs including M06-2X or SCC-DF-TB-D not are available. Overlays of structures optimized using either B3LYP-DCP or RI-MP2/cc-PVTZ are displayed in the Supporting Information.

In general terms, all 15 isomers except isomer **9** feature a hydrogen bond between the amide hydrogen of Gly and the N-terminal amino group (Fig. 6). The hydrogen bond is ranging from 2.10 to 2.21 Å forming a five-membered ring (ring **I**).

The strength of this hydrogen bond is estimated by the energy difference between isomers **1** and **9** to be 2.3 kcal mol$^{-1}$. As shown from the overlay of the two structures in Fig. 7, the only characteristic difference between them is the rotation of the amino group, which makes the nitrogen lone pair point away from the possibility of forming a NH···NH$_2$ hydrogen bond.

In isomers **2**, **7** and **12**, the amino group is rotated about the C$_{18}$–N bond so that its lone-pair is delocalized via a negative hyperconjugation out to the σ*–C$_{18}$H bond [40]. Notably, isomer **1**, which is the most stable isomer, and isomer **14**, which is the least stable, both show the strongest delocalization of this nitrogen lone-pair

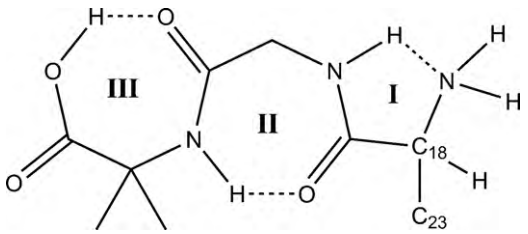


**Fig. 6.** Portion of the tripeptide Phe-Gly-Phe system.

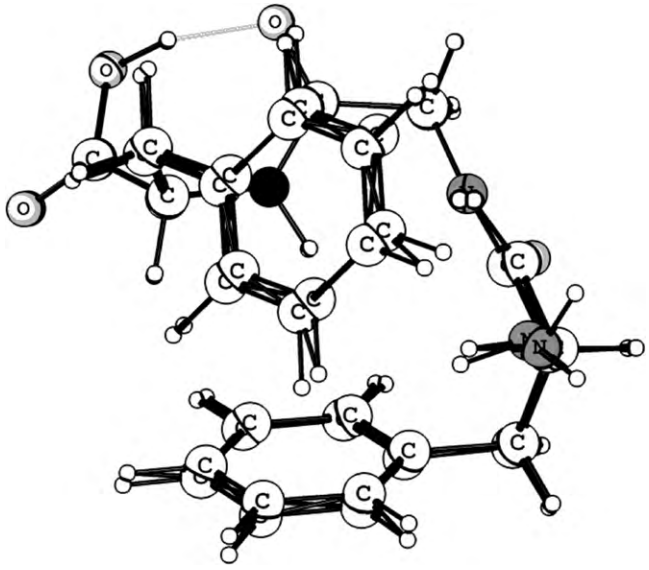

**Fig. 7.** Overlay of isomers **1** and **9**.

to the nearby σ*(C–C) orbital, as evidenced from the elongated C$_{18}$–C$_{23}$ bond distance of 1.548 Å. Only isomer **15**, which also is among the least stable isomers, shows a longer C–C bond (1.553 Å). Thus, there is no direct correlation between the negative hyperconjugation effect and relative energy, showing that many components are involved in determining the overall relative energies among the different isomers. Out of the 15 isomers studied, ten form a cyclic seven-membered central ring (ring **II**, Fig. 6) due to a hydrogen bond between the amide hydrogen from the C-terminal Phe residue and the carbonyl oxygen of Gly (column 2, Table S2). The distance in the hydrogen bond varies from 1.89 to 2.04 Å. Of all seven S- or T-stacked isomers all except one (isomer **8**) contains this hydrogen bond. Another seven-membered ring can be formed via a hydrogen bond (column 3, Table S2) between the carboxyl group and the carbonyl oxygen of the glycine (ring **III**, Fig. 6). An estimation of this hydrogen bond strength is taken from the energy difference (2.9 kcal mol$^{-1}$) between isomers **1** and **14**. An overlay of these two structures reveals that the carboxyl group is rotated in the latter, removing the possibility of forming a hydrogen bond with the carbonyl oxygen. The strength of this hydrogen bond is estimated to be stronger than the NH···NH$_2$ hydrogen bond discussed above. This seems reasonable in view of the shorter OH···O hydrogen bond (ca. 1.7 Å vs. ca. 2.1 Å) and the larger electronegativity of oxygen compared to nitrogen. Eight of the 15 isomers do not form ring **III** and out of these, five are from the group having a phenyl–phenyl interaction. Thus, it appears that the S- and T-stacked isomers of Phe-Gly-Phe preferably have a NH···OC hydrogen bond interaction but no COOH···OC hydrogen bond, although the statistical base for this is rather weak. There is no correlation found between hydrogen bond distances (or sums thereof) and relative energies, emphasizing the fact that many small interactions are delicately balanced to give the relative stabilities of the different isomers. Valdes and co-workers draw the conclusion that the maximum *number of intramolecular interactions* would lead to the most stable isomer. However, as seen from Table S2, where a summary of interactions are given, simply counting the number of hydrogen bonds, aromatic-aromatic interactions, and XH–π (X = C, N) interactions does not directly correlate with the total stability since the strength of these interactions are varying dependent on the interacting distances. In addition, repulsive interactions need to be included in such a summation to get the overall stability.

In conclusion, the B3LYP-DCP study on the tripeptide Phe-Gly-Phe has shown that the method generates optimized geometries

in good agreement with more demanding MP2-optimizations. In addition, it has been shown to give accurate relative energies of different isomers of peptide systems where hydrogen bonds, aromatic-aromatic interactions, and XH–π (X = C, N) interactions are competitive. Mean absolute deviations from calculated CCSD(T)/CBS//RI-MP2/cc-pVTZ relative energies is in the order of 0.50 kcal mol$^{-1}$ and for complexes including stacked or T-oriented aromatic rings even smaller.

### 3.2. Phenylalanine derivative

Evidence of weak interactions in chemical systems is rather difficult to obtain experimentally. Edge to face interactions between aromatic rings, such as CH–π interactions are notoriously difficult to characterize accurately with for example NMR spectroscopy due to fast dynamical processes such as ring or alkyl group rotations. In the phenylalanine derivative **16**, phenyl ring rotation is partially restricted due to the incorporation of an imine functionality (Fig. 8) [29]. This retardation of dynamics allows for separation of the $^1$H NMR signals of the two ortho protons. One of the ortho protons will experience an upfield shift when positioned above the plane of the phenol ring due to a ring current effect, while the other will show a signal typical for aromatic protons (ca. 7–7.5 ppm).

In addition to NMR measurements, Jennings et al. were able to grow crystals of compound **16** and generate an X-ray structure [29]. Comparison of some key bond distances is therefore useful to evaluate the B3LYP-DCP method. Most importantly, in the B3LYP-DCP optimized structure it is observed that the two aromatic rings do interact as observed in the crystal structure (Fig. 9A).

The distance from the H-ortho' proton to the ring centre of the phenol ring situated below is 2.63 Å compared to 2.60 Å in the crystal structure. Similarly, the perpendicular distance from the H-ortho' proton to the ring plane is 2.61 Å compared to 2.59 Å in the X-ray structure. Thus, the structure of **16** is very well described by the B3LYP-DCP method. In Fig. 10, the CH–π interaction between the H-ortho' proton and the aromatic ring is visualized in terms of NBO-overlap (π–σ*(CH) interaction).

In contrast, removing the DCP term and re-optimizing compound **16** generates a structure without CH–π interaction. Now, the distance between the H-ortho' proton and the ring centre is 3.75 Å in comparison to 2.60 Å in the crystal structure (Fig. 9B). This clearly demonstrates the importance of including the DCP for describing the structure accurately.

Experimentally, Jennings et al. found that the difference in chemical shift between the two ortho protons was as large as 1.36 ppm (7.14–5.78 ppm) [29]. Employing B3LYP/6-31+G(d,p)-DCP calculations it is found that the difference in chemical shift is 2.10 ppm (7.25–5.15 ppm). Thus, B3LYP-DCP accurately predicts the upfield shift of the H-ortho' proton. The overestimated shift difference may be due to the fact that experimentally not only one conformer of **16** is observed, and thus the true difference is "diluted"

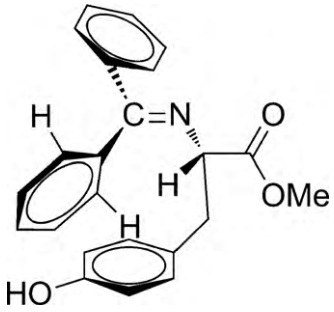


**Fig. 8.** Compound **16**.

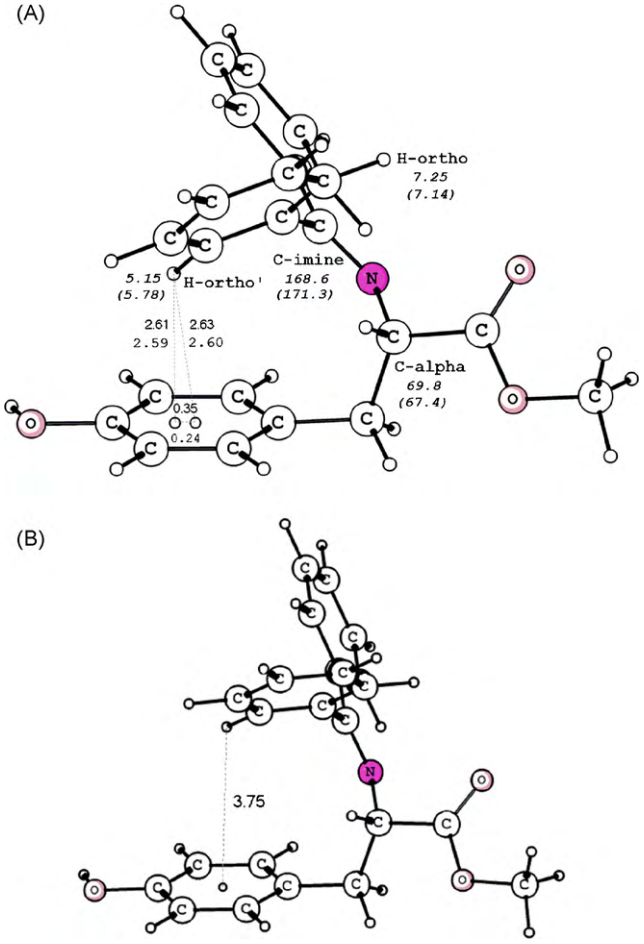


**Fig. 9.** (A) B3LYP/6-31+G(d,p)-DCP optimized compound **16**. (B) B3LYP/6-31+G(d,p) optimized compound **16**. Bond distances (in Å) in bold are compared with bond distances (in normal) from the crystal structure. Selected chemical shieldings (in ppm) are reported compared to experimentally determined shifts (in parenthesis).

by other smaller differences observed in other conformers populated. From structural data (*vide supra*), Jennings et al. estimated the difference between the two ortho protons to be ca. 2.6 ppm, which correlates better with the B3LYP-DCP results. To control if the calculated chemical shifts were sensitive to the inclusion of

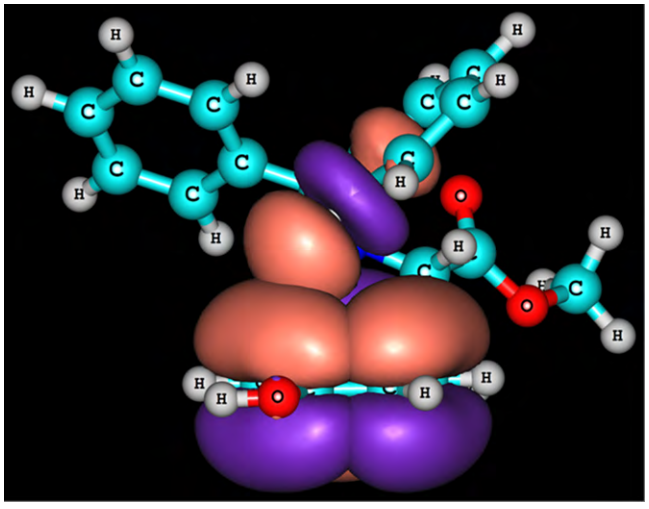

**Fig. 10.** π–σ*(CH) interaction in compound **16**.

the DCP, calculations without employing the DCP in the NMR calculation (B3LYP/6-31+G(d,p)//B3LYP/6-31+G(d,p)-DCP) were also performed. Only a negligible difference (7.25–5.16 ppm) to those reported above was experienced. This informs that inclusion of the DCP does not influence the accuracy of the actual NMR calculation. As a second control, NMR chemical shieldings were also calculated using the triple zeta basis set 6-311+G(d,p) to see if this significantly would improve the results or alter the conclusion. With the larger basis set it was found that the shift for H-ortho' proton was slightly less upfield shifted (5.28 ppm) and thus in better agreement with the experimental data (5.78 ppm). On the other hand, the chemical shift for the second ortho proton was in slightly worse agreement with experiment than that found using the smaller double zeta basis set. The difference between the two ortho protons calculated using 6-311+G(d,p) was 2.17 ppm, which is slightly larger than that found using 6-31+G(d,p). Overall, there seems to be small effects from the basis set, and negligible effect from the DCP, on the chemical shifts for compound **16**. For the structure optimized without the DCP correction, the lack of influence from the ring current on the H-ortho' proton makes it resonate at a less upfield shift (6.34 ppm) compared to that calculated for the B3LYP-DCP optimized structure.

### 3.3. The dipeptide Ac-Phe-Phe-$NH_2$

The last system in this investigation was studied in the gas phase by Gloaguen et al. using IR- and UV-spectroscopy in combination with computational methodology [30]. Assignment of experimental data excluded all but two different conformers of the dipeptide, the g+g+ and g−g+ isomers (Fig. 11).

Gloaguen et al. found that an edge-to-face interaction between the two aromatic rings was crucial in stabilization of the most sta-

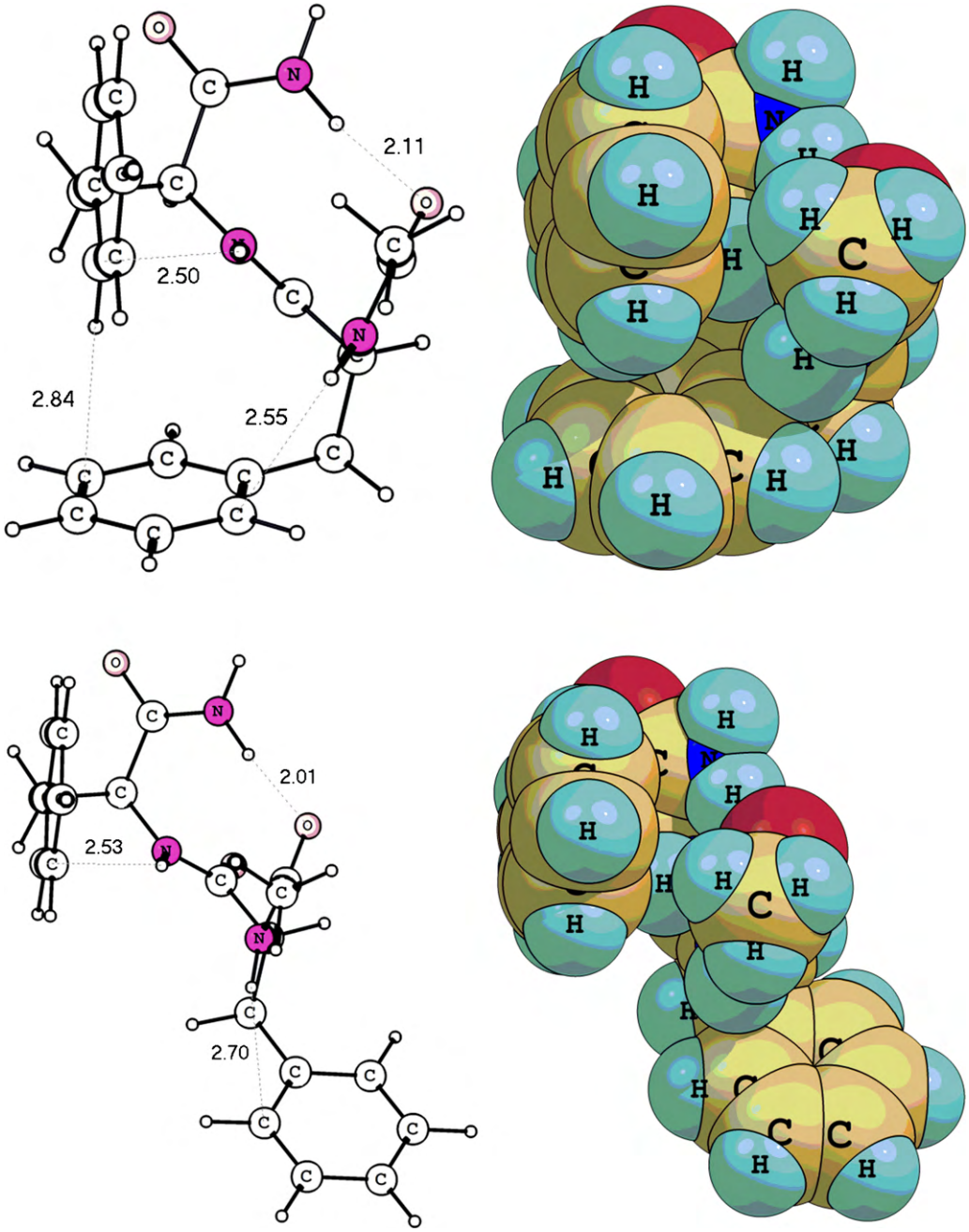


**Fig. 11.** B3LYP/6-31+G(d,p)-DCP optimized structures of the g+g+ (top) and g−g+ (bottom) conformers of the Ac-Phe-Phe-$NH_2$ dipeptide.

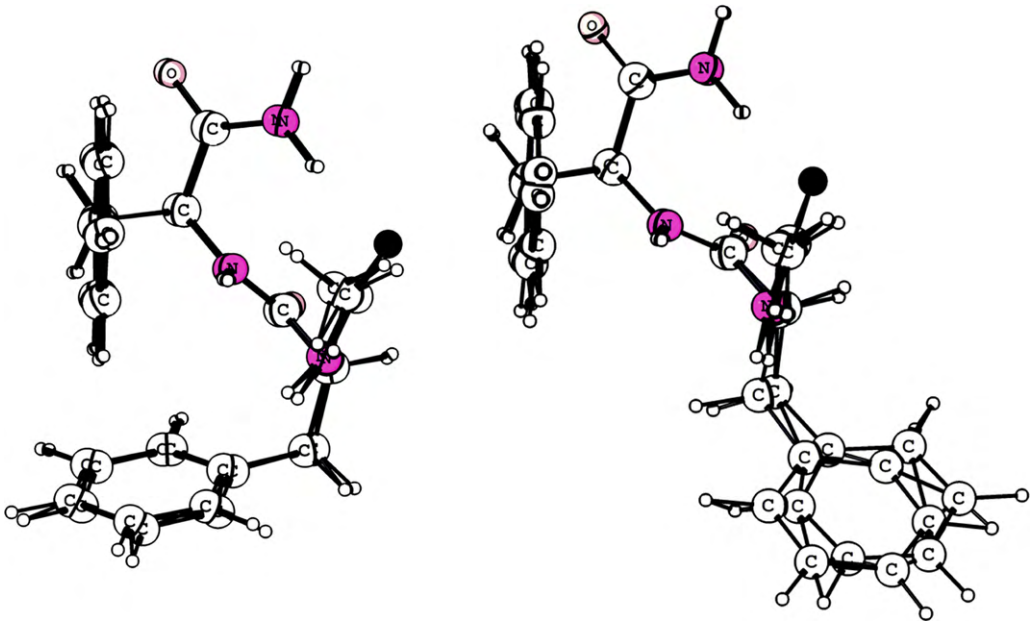

**Fig. 12.** Overlay of RI-B97-D/TZVPP and B3LYP-DCP optimized structures of the g+g+ (left) and g−g+ (right) isomers of the Ac-Phe-Phe-$NH_2$ dipeptide.

ble g+g+ isomer. This interaction competes with hydrogen bonds and NH–π interactions as deduced from the optimized structures. It was found that the computational methods used in their study were good enough to describe the structures and relative energies of the two conformers. However, the assignment based on calculated frequencies was not straight-forward and only in combination with energy data could the correct isomer be identified.

Subtle differences in for example NH–π interactions and $NH_2\cdots O{=}C-$ hydrogen bonds will affect the vibrational frequencies and these are therefore useful and sensitive structural probes to use for identification of specific interactions. A good agreement between calculated and experimental vibrational frequencies is therefore likely an indication that the correct structure has been identified. However, one need to be precautious in the interpretation; Gloaguen et al. showed that both B3LYP/6-31+G(d) and RI-SCS-MP2/SVP calculations gave a better agreement between calculated and experimental frequencies for the incorrect, less stable g−g+ isomer. We therefore wanted to employ the B3LYP-DCP method to see if this could aid in the direct assignment of experimental vibrational frequencies to a specific conformer of the dipeptide Ac-Phe-Phe-$NH_2$.

In the study by Gloaguen et al. it was found that RI-B97-D/TZVPP calculations resulted in an energy difference of 3.35 kcal $mol^{-1}$ between the g+g+ and g−g+ isomers. The corresponding energy difference using B3LYP/6-31+G(d,p)-DCP is 2.24 kcal $mol^{-1}$, giving a deviation of 1.1 kcal $mol^{-1}$. We saw from the study of the tripeptide and the phenylalanine derivative that structures optimized using B3LYP-DCP are in excellent agreement with experimental data or other calculations. This is verified by overlaying the structures of the g+g+ and g−g+ isomers optimized with RI-B97-D/TZVPP or B3LYP/6-31+G(d,p)-DCP (Fig. 12).

The RMSD values for the structures are 0.09 and 0.24 Å for the g+g+ and g−g+ isomers, respectively. The higher value for the latter isomer comes from the deviation in the exact location of the remote phenyl ring. This is evidenced by examining the distances between the centres of the two aromatic rings in the isomers. In the RI-B97-D/TZVPP optimized structures the distances are 5.12 and 7.94 Å, respectively, while in the B3LYP/6-31+G(d,p)-DCP optimized structures they are 5.25 and 8.32 Å, respectively. Besides the edge-to-face interaction between the two aromatic rings the hydrogen bond pattern of the dipeptide is also well described. For example, the $NH_2$–O hydrogen bond is calculated to be 2.11 and 2.01 Å in the g+g+ and g−g+ isomers, respectively (Fig. 11),

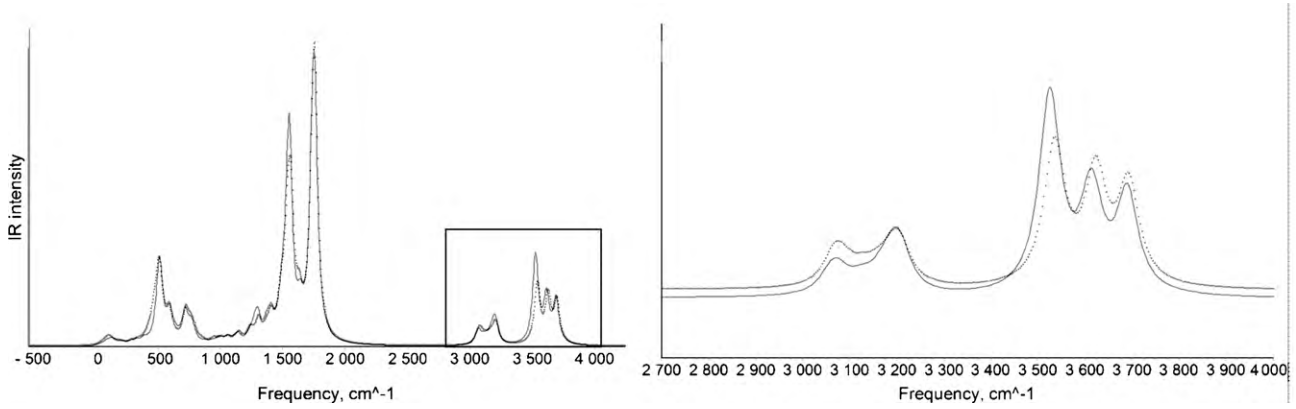


**Fig. 13.** B3LYP-DCP calculated IR spectra for the g+g+ (dashed line) and g−g+ (solid line) isomer of the Ac-Phe-Phe-$NH_2$ dipeptide. The boxed area is enlarged in the right portion of the figure.

**Table 3**
Experimental and calculated vibrational frequencies (in $cm^{-1}$).[a].

| | Isomer | Scaling factor | $NH_2$ (symm.) | NH (1) | NH (2) | $NH_2$ (asymm.) | RMSD |
|---|---|---|---|---|---|---|---|
| Exp. | | | 3391 | 3437 | 3447 | 3524 | |
| B3LYP-DCP (unscaled) | g+g+ | – | 3536 (+145) | 3621 (+184) | 3626 (+179) | 3694 (+170) | 170.3 |
| | g−g+ | – | 3525 (+134) | 3611 (+174) | 3617 (+170) | 3690 (+166) | 161.9 |
| B3LYP-DCP (uniform scaling) | g+g+ | 0.9531 | 3370 (−21) | 3452 (+15) | 3456 (+9) | 3520 (−4) | 13.5 |
| | g−g+ | 0.9553 | 3368 (−23) | 3450 (+13) | 3455 (+8) | 3525 (+1) | 13.8 |
| B3LYP-DCP (mode separated scaling) | g+g+ | 0.9547/0.9565/0.9515 | 3376 (−15) | 3464 (+27) | 3469 (+22) | 3514 (−10) | 19.4 |
| | g−g+ | 0.9547/0.9565/0.9515 | 3366 (−25) | 3454 (+17) | 3459 (+12) | 3511 (−13) | 17.6 |

[a] Deviations from the experimental values are given in parenthesis.

compared to 2.20 and 2.04 Å in the RI-B97-D/TZVPP optimized structures. Thus, the elongation of the hydrogen bond in the g+g+ isomer is accurately reproduced using B3LYP-DCP. As can be seen in Fig. 13, the two calculated spectra for the isomers are very similar with only small deviations in the high-frequency region where the $NH_2$ and NH stretches are found. The difference in hydrogen bond strength is indeed seen in the change in the $NH_2$ stretches, which in the g−g+ isomer shows a lower frequency (red shifted) than in the g+g+ isomer according to the B3LYP-DCP calculations (Table 3 and Fig. 13). However, the frequency differences are only in the order of 5–10 $cm^{-1}$, and thus very challenging to interpret. The unscaled B3LYP-DCP calculated harmonic frequencies show a rather large deviation from the experimental values (RMSD 162–170 $cm^{-1}$). This is a typical feature of calculated frequencies and is corrected for by introducing a method and basis set dependent scaling factor. For example, Scott and Radom were among the first to systematically derive scaling factors for a variety of *ab initio* and DFT methods [41]. The smallest deviation achieved for the B3LYP/6-31+G(d,p)-DCP calculated $NH_2$ and NH vibrational frequencies was found when applying scaling factors 0.9531 for the g+g+ isomer, and 0.9553 for the g−g+ isomer. Now, the RMS deviations from the experimental values could be reduced to 13.5 and 13.8 $cm^{-1}$, respectively, but with maximum errors as large as 21 and 23 $cm^{-1}$, respectively. The negligible difference in accuracy comparing the two isomers and the magnitude of the error thus does not allow for assigning the experimental frequencies to either the g+g+ or the g−g+ isomer. An alternative scaling was also tested, with separate scaling factors for the $NH_2$ asymmetric, $NH_2$ symmetric and NH stretches. These scaling factors were derived by comparing experimental frequencies for acetamide and *N*-methylacetamide with B3LYP/6-31+G(d,p)-DCP calculated frequencies. However, as seen in Table 3, the agreement between calculated and experimental frequencies with these mode separated scaling factors was found in less satisfying agreement than using the uniform scaling factor. A similar observation was found by Gloaguen and co-workers. An alternative solution might be to scale the force constants rather than the frequencies [42], but this has not been applied in this study. The conclusion from the results presented above, is thus that the B3LYP/6-31+G(d,p)-DCP calculated geometries are in good agreement with those obtained using other dispersion-corrected DFT methods. However, the accuracy of the calculated vibrational frequencies does not fully allow separating isomers with small (<10 $cm^{-1}$) differences in vibrational frequencies.

## 4. Conclusion

An evaluation of a dispersion-corrected DFT method (B3LYP-DCP) for three systems of biochemical interest has been performed in order to increase the knowledge on the general applicability of the method. The first system investigated, the tripeptide Phe-Gly-Phe, was selected in order to compare B3LYP-DCP and benchmark CCSD(T)/CBS calculated relative energies of 15 different peptide isomers. In the different isomers, different weak interactions such as aromatic ring–ring interactions, CH–π- and NH–π interactions, or hydrogen bonds all compete in forming the most stable structure. The balance between the different interactions is found to be well described using B3LYP-DCP since the mean absolute deviation in the relative energies compared to the CCSD(T)/CBS calculations was found to be only 0.50 kcal $mol^{-1}$. It was also found that both B3LYP-DCP and CCSD(T)/CBS describe the same isomer as the most stable. In addition, B3LYP-DCP and RI-MP2 optimized structures have been compared showing on a good agreement (RMSD 0.15 Å). In the second system, a phenylalanine derivative was investigated. Here, the comparison of the B3LYP-DCP optimized structure and X-ray crystal data nicely showed that the dispersion-corrected DFT method accurately predicts the CH–π interaction between the two aromatic rings. This interaction was computationally verified by a strongly upfield shift for one of the protons in the aromatic ring in agreement with experimental $^1H$ NMR data. In the third system covered in the present study an edge-to-face interaction between two aromatic rings was found well described using B3LYP-DCP. Attempts to use B3LYP-DCP calculated vibrational frequencies in the assignment of experimental IR-data for this system did not allow separation of the two most stable isomers. The differences in calculated frequencies (<10 $cm^{-1}$) were of the same magnitude as the relative error compared to the experimental data, and thus only in combination with energy calculations could the correct isomer be predicted.

The present study shows that B3LYP-DCP is capable of describing geometries and energies of compounds which feature interactions typically observed in compounds of biochemical interest. Thus, the method could for example be readily used for accurate studies of interactions between drugs and an active site in a protein, or for studying reaction mechanisms catalyzed by enzymes. In combination with previous studies using the DFT-DCP approach [12–18], the method has now shown on its general applicability.

## Acknowledgement

S.O.N.L. gratefully acknowledges a post-doctoral stipend from the Sven and Lilly Lawski Foundation. Dr. Eric Gloaguen is acknowledged for kindly providing the coordinates for the g+g+ and g-g+ isomers.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmgm.2010.06.001.

## References

[1] D. Fattori, Molecular recognition: the fragment approach in lead generation, Drug Discov. Today 9 (2004) 229–238.
[2] G. Kryger, I. Silman, J.L. Sussman, Structure of acetylcholinesterase complexed with E2020 (Aricept®): implications for the design of new anti-Alzheimer drugs, Structure 7 (1999) 297–307.
[3] S. Burley, G. Petsko, Aromatic–aromatic interaction: a mechanism of protein structure stabilization, Science 229 (1985) 23–28.

[4] T. Ishida, M. Doi, H. Ueda, M. Inoue, G.M. Scheldrick, Specific ring stacking interaction on the tryptophan-7-methylguanine system: comparative crystallographic studies of indole derivatives-7-methylguanine base, nucleoside, and nucleotide complexes, J. Am. Chem. Soc. 110 (1988) 2286–2294.
[5] G.B. McGaughey, M. Gagné, A.K. Rappé, Pi-stacking interactions, J. Biol. Chem. 273 (1998) 15458–15463.
[6] C. Janiak, A critical account on pi–pi stacking in metal complexes with aromatic nitrogen-containing ligands, J. Chem. Soc., Dalton Trans. (2000) 3885–3896.
[7] C.A. Hunter, K.R. Lawson, J. Perkins, C.J. Urch, Aromatic interactions, J. Chem. Soc., Perkin Trans. 2 (2001) 651–669.
[8] R. Bhattacharyya, U. Samanta, P. Chakrabarti, Aromatic–aromatic interactions in and around {alpha}-helices, Protein Eng. 15 (2002) 91–100.
[9] A. Thomas, R. Meurisse, R. Brasseur, Aromatic side-chain interactions in proteins. II. Near- and far-sequence Phe-X pairs, Proteins: Struct. Funct. Genet. 48 (2002) 635–644.
[10] E.A. Meyer, R.K. Castellano, F. Diederich, Interactions with aromatic rings in chemical and biological recognition, Angew. Chem., Int. Ed. 42 (2003) 1210–1250.
[11] M. Nishio, Y. Umezawa, K. Honda, S. Tsuboyama, H. Suezawa, CH/π hydrogen bonds in organic and organometallic chemistry, Cryst. Eng. Commun. 11 (2009) 1757–1788.
[12] G. DiLabio, Accurate treatment of van der Waals interactions using standard density functional theory methods with effective core-type potentials: application to carbon-containing dimers, Chem. Phys. Lett. 455 (2008) 348–353.
[13] I.D. Mackie, G.A. DiLabio, Interactions in large, polyaromatic hydrocarbon dimers: application of density functional theory with dispersion corrections, J. Phys. Chem. A 112 (2008) 10968–10976.
[14] S.O. Nilsson Lill, Application of dispersion-corrected density functional theory, J. Phys. Chem. A 113 (2009) 10321–10326.
[15] I.D. Mackie, S.A. McClure, G.A. DiLabio, Binding in thiophene and benzothiophene dimers investigated by density functional theory with dispersion-correcting potentials, J. Phys. Chem. A 113 (2009) 5476–5484.
[16] G.A. DiLabio, E.R. Johnson, J. Pitters, Pentacene binds strongly to hydrogen-terminated silicon surfaces via dispersion interactions, J. Phys. Chem. C 113 (2009) 9969–9973.
[17] E.R. Johnson, G.A. DiLabio, Theoretical study of dispersion binding of hydrocarbon molecules to hydrogen-terminated silicon(1 0 0)-2 × 1, J. Phys. Chem. C 113 (2009) 5681–5689.
[18] G.A. Shamov, P.H.M. Budzelaar, G. Schreckenbach, Performance of the empirical dispersion corrections to density functional theory: thermodynamics of hydrocarbon isomerizations and olefin monomer insertion reactions, J. Chem. Theory Comput. 6 (2010) 477–490.
[19] S. Grimme, Accurate description of van der Waals complexes by density functional theory including empirical corrections, J. Comput. Chem. 25 (2004) 1463–1473.
[20] S. Grimme, Semiempirical GGA-type density functional constructed with a long-range dispersion correction, J. Comput. Chem. 27 (2006) 1787–1799.
[21] S. Grimme, J. Antony, T. Schwabe, C. Mück-Lichtenfeld, Density functional theory with dispersion corrections for supramolecular structures, aggregates, and complexes of (bio)organic molecules, Org. Biomol. Chem. 5 (2007) 741–758.
[22] Y. Zhao, D. Truhlar, The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals, Theor. Chem. Acc. 120 (2008) 215–241.
[23] M. Dion, H. Rydberg, E. Schröder, D.C. Langreth, B.I. Lundqvist, Van der Waals density functional for general geometries, Phys. Rev. Lett. 92 (2004) 246401.
[24] J. Gräfenstein, D. Cremer, An efficient algorithm for the density-functional theory treatment of dispersion interactions, J. Chem. Phys. 130 (2009) 124105.
[25] L.A. Curtiss, K. Raghavachari, P.C. Redfern, J.A. Pople, Investigation of the use of B3LYP zero-point energies and geometries in the calculation of enthalpies of formation, Chem. Phys. Lett. 270 (1997) 419–426.
[26] S.F. Sousa, P.A. Fernandes, M.J. Ramos, General performance of density functionals, J. Phys. Chem. A 111 (2007) 10439–10452.
[27] P. Jurecka, J. Sponer, J. Cerný, P. Hobza, Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs, Phys. Chem. Chem. Phys. 8 (2006) 1985–1993.
[28] H. Valdes, K. Pluhackova, P. Hobza, Phenylalanyl-glycyl-phenylalanine tripeptide: a model system for aromatic–aromatic side chain interactions in proteins, J. Chem. Theory Comput. 5 (2009) 2248–2256.
[29] W.B. Jennings, N.J.P. McCarthy, P. Kelly, J.F. Malone, Topically resolved intramolecular CH–π interactions in phenylalanine derivatives, Org. Biomol. Chem. 7 (2009) 5156–5162.
[30] E. Gloaguen, H. Valdes, F. Pagliarulo, R. Pollet, B. Tardivel, P. Hobza, F. Piuzzi, M. Mons, Experimental and theoretical investigation of the aromatic–aromatic interaction in isolated capped dipeptides, J. Phys. Chem. A 114 (2009) 2973–2982.
[31] A.D. Becke, Density-functional thermochemistry. III. The role of exact exchange, J. Chem. Phys. 98 (1993) 5648–5652.
[32] C. Lee, W. Yang, R.G. Parr, Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density, Phys. Rev. B: Condens. Matter 37 (1988) 785–789.
[33] P.J. Stephens, F.J. Devlin, C.F. Chabalowski, M.J. Frisch, Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields, J. Phys. Chem. 98 (1994) 11623–11627.
[34] W.J. Hehre, R. Ditchfield, J.A. Pople, Self-consistent molecular orbital methods. XII. Further extensions of Gaussian-type basis sets for use in molecular orbital studies of organic molecules, J. Chem. Phys. 56 (1972) 2257–2261.
[35] R. Krishnan, J.S. Binkley, R. Seeger, J.A. Pople, Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions, J. Chem. Phys. 72 (1980) 650–654.
[36] G.W. Spitznagel, T. Clark, J. Chandrasekhar, P.v.R. Schleyer, Stabilization of methyl anions by first-row substituents. The superiority of diffuse function-augmented basis sets for anion calculations, J. Comput. Chem. 3 (1982) 363–371.
[37] G.W. Frisch, 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.W. 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. Johnson, W. Chen, M.W. Wong, C. Gonzalez, J.A. Pople, Gaussian 03. Revision D.02, Gaussian Inc., Wallingford, CT, 2004.
[38] K. Wolinski, J.F. Hinton, P. Pulay, Efficient implementation of the gauge-independent atomic orbital method for NMR chemical shift calculations, J. Am. Chem. Soc. 112 (1990) 8251–8260.
[39] G.A. Zhurko, http://www.chemcraftprog.com, 2009.
[40] S.O. Nilsson Lill, G. Rauhut, E. Anders, Chemical reactivity controlled by negative hyperconjugation: a theoretical study, Chem. A: Eur. J. 9 (2003) 3143–3153.
[41] A.P. Scott, L. Radom, Harmonic vibrational frequencies: an evaluation of Hartree–Fock, Moeller–Plesset, quadratic configuration interaction, density functional theory, and semiempirical scale factors, J. Phys. Chem. 100 (1996) 16502–16513.
[42] G. Rauhut, P. Pulay, Transferable scaling factors for density functional derived vibrational force fields, J. Phys. Chem. 99 (1995) 3093–3100.