

# **Intramolecular BSSE and dispersion affect the structure of a dipeptide conformer**

Rabia Hameed<sup>a</sup>, Afsar Khan<sup>b</sup>, and Tanja van Mourik<sup>c\*</sup>

<sup>a</sup> *EaStCHEM School of Chemistry, University of St Andrews, North Haugh, St Andrews KY16 9ST, United Kingdom. Current address: Department of Chemistry, COMSATS Institute of Information Technology Abbottabad 22060, Pakistan*Department, University, City, Country

<sup>b</sup> *Department of Chemistry, COMSATS Institute of Information Technology, Abbottabad 22060, Pakistan*

<sup>c</sup> *EaStCHEM School of Chemistry, University of St Andrews, North Haugh, St Andrews KY16 9ST, United Kingdom. Corresponding author. Email: [tanja.vanmourik@st-andrews.ac.uk](mailto:tanja.vanmourik@st-andrews.ac.uk). Phone: +44 1334 463822.*

## **Intramolecular BSSE and dispersion affect the structure of a dipeptide conformer**

B3LYP and MP2 calculations with the commonly-used 6-31+G(d) basis set predict qualitatively different structures for the Tyr-Gly conformer book1, which is the most stable conformer identified in a previous study (Mol. Phys. 104, 559-570, 2006). The structures differ mainly in the  $\psi_{\text{Tyr}}$  Ramachandran angle ( $138^\circ$  in the B3LYP structure and  $120^\circ$  in the MP2 structure). The causes for the discrepant structures are attributed to missing dispersion in the B3LYP calculations and large intramolecular BSSE in the MP2 calculations. The correct  $\psi_{\text{Tyr}}$  value is estimated to be  $130^\circ$ . The MP2/6-31+G(d) profile identified an additional conformer, not present on the B3LYP surface, with a  $\psi_{\text{Tyr}}$  value of  $96^\circ$  and a more folded structure. This minimum is however likely an artefact of large intramolecular BSSE values. We recommend the use of basis sets of at least quadruple-zeta quality in DFT, DFT-D and MP2 calculations in cases where intramolecular BSSE is expected to be large.

Keywords: tyrosine-glycine, basis set superposition error, density functional theory, MP2, DFT-D

### **Introduction**

Conformational analysis of peptides is a challenging problem. Because of their inherent flexibility, even small peptides have very many possible conformers. One common strategy to identify the most stable conformer is to search the conformational space with an hierarchical methodology: the large pool of potential conformers is treated at a fast, low-level of theory, and the most stable conformers according to this low level of theory are re-optimised at increasingly higher levels of theory [1-8]. In our group, we have used such a hierarchical selection method to study small peptides including Tyr-Gly (Tyr = tyrosine; Gly = glycine) [5], Tyr-Gly-Gly [9] and Gly tripeptide [10]. The methodology used for these peptides included a combination of Hartree-Fock (HF), density functional theory (DFT) employing the B3LYP [11,12] functional and second-

order Møller-Plesset (MP2) perturbation theory. However, such hierarchical methodologies risk missing conformers because of deficiencies in the lower-level methods employed. Indeed, even the ‘higher-level’ MP2 and B3LYP methods may not be accurate enough, particularly when used with limited basis sets. For the Tyr-Gly dipeptide, we found several instances where B3LYP/6-31+G(d) and MP2/6-31+G(d) geometry optimisation gave strikingly different results [5]. In general, MP2 gives more ‘folded’, compact, conformers. There are two possible reasons for this. Firstly, MP2 calculations with small to medium-sized basis sets are plagued by intramolecular basis set superposition error (BSSE) effects, which cause an artificial attraction between different parts of the molecule. BSSE tends to be smaller in DFT calculations. Secondly, functionals like B3LYP do not describe London dispersion forces, which are attractive. These two effects, dispersion (attractive) and BSSE (also attractive, but artificial) are often large in molecules containing an aromatic ring.

Three examples of discrepant Tyr-Gly structures found by B3LYP/6-31+G(d) and MP2/6-31+G(d) geometry optimisations are shown in Figure 1. Book1 was identified as the most stable Tyr-Gly conformer in Ref. [1]. The other two examples (book4 and book6) are among the ten most stable conformers identified in that study. The numbering of the conformers follows stability according to MP2/6-31+G(d)//B3LYP/6-31+G(d) calculations (increasing numbers denote decreasing stability); the ‘book’ attribute denotes these conformers have a folded structure. The MP2 and B3LYP book4 structures mainly differ in the value of the  $\phi_{\text{Gly}}$  Ramachandran dihedral angle, which is  $180^\circ$  in the B3LYP minimum and  $74^\circ$  in the MP2 minimum. Calculation of energy profiles for variation of this dihedral angle revealed two minima in the MP2 profile, at  $74^\circ$  and  $\sim 280^\circ$ , and only one minimum in the B3LYP profile, at  $180^\circ$  [2]. It was found that large intramolecular BSSE values hide the  $180^\circ$ -minimum in

the MP2 profile (which only shows up after BSSE correction or reduction), whereas B3LYP misses the minima at  $74^\circ$  and  $\sim 280^\circ$  presumably due to lacking dispersion. A further study showed that density functionals that are capable of describing dispersion, including the Minnesota functionals M05-2X [3], M06-2X [4] and M06-L [5], the double hybrid functional mPW2-PLYP [6], and DFT-D [7,8] methods (B3LYP-D and mPW2-PLYP-D) found all three minima [9], thereby supporting this presumption. For book6, MP2/6-31+G(d) predicts a ‘closed book’ conformation, whereas B3LYP/6-31+G(d) predicts a more open conformation. A further study found that the closed-book conformer is an artefact caused by large BSSE values in the MP2 calculations [10].

In the current work we investigate book1. For this conformer, the B3LYP and MP2 optimised structures mainly differ in the  $\psi_{\text{tyr}}$  Ramachandran angle, which is  $138^\circ$  in the B3LYP structure and  $120^\circ$  in the MP2 structure.

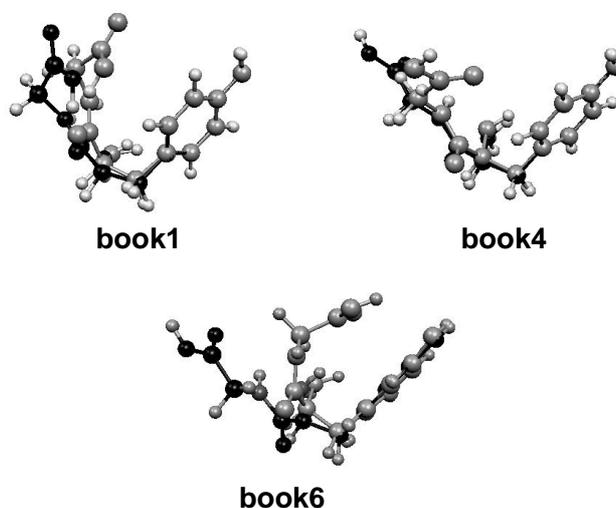


Figure 1. Comparison of the B3LYP/6-31+G(d) (black atoms) and MP2/6-31+G(d) (grey atoms) geometries of Tyr-Gly conformers book1, book4 and book6.

Intermolecular BSSE can be eliminated with the counterpoise (CP) method [11], however, there is no unambiguous way to eliminate *intramolecular* BSSE. Some

approximate schemes have been proposed. Most of these are based on the CP scheme, though Jensen recently introduced a non-CP approach based on valence bond theory [12]. Palermo *et al.* proposed a method to correct for intramolecular BSSE between an aromatic residue and the backbone in peptides [13]. This method involves rotating the aromatic residue around a suitable bond to create a ‘non-interacting’ conformation and adding in both (interacting and non-interacting) conformations ghost functions at the position of the other conformation. However, it has been shown that this scheme underestimates the magnitude of the BSSE [14]. A revised rotation method has subsequently been proposed [15]. The rotation method relies on the availability of a suitable bond to ‘rotate apart’ the two interacting parts of the molecule and is not readily applicable to rotational energy profiles. A method to correct for both inter- and intramolecular BSSE in calculations on large molecules using small basis sets was presented by Kruse and Grimme [16]. This semi-empirical method, dubbed gCP, only depends on the geometry of the system of interest, with no input from the electronic wavefunction. We have previously used a fragment-based method to estimate the BSSE in Tyr-Gly [2,10,17]. This fragmentation method estimates *intramolecular* BSSE for the molecule of interest from *intermolecular* BSSE calculated with the CP method for separate fragments modelling the original molecule [17]. In the fragmentation method, the molecule is split into the two interacting fragments; some atoms in the linkage are removed to avoid overlapping atoms and the dangling bonds of the two fragments are saturated with hydrogens. The CP method is then employed to calculate the BSSE between the two interacting fragments. Note there is some ambiguity regarding the choice of fragments in this method. Other fragment-based approaches have been proposed: Asturiol *et al.* used a CP scheme using intramolecular fragments such as C-H, C=O, N-H and CCH<sub>3</sub> to show that the observed non-planarity of aromatic molecules,

including benzene and the nucleobases, is attributable to intramolecular BSSE [18,19]. Balabin used a similar scheme to study intramolecular BSSE in alkanes [20,21]. Jensen proposed an atomic counterpoise method that can be applied to estimate inter- as well as intramolecular BSSE. It estimates the BSSE as a sum of atomic contributions [22].

BSSE can be reduced by increasing the basis set size. Though it is generally observed that BSSE decreases with increasing basis set size, this is not always the case. For example, for He<sub>2</sub> the BSSE in CCSD(T) calculations increases from the aug-cc-pVnZ to d-aug-cc-pVnZ to t-aug-cc-pVnZ (n = D, T, Q, 5 and 6) basis set series [23], presumably because the more diffuse basis sets offer more opportunity for one He atom to use the basis functions of the other. Likewise, it had been observed in earlier work that basis functions optimised on the dispersion energy enhance the magnitude of BSSE [24]. In the case of the x-aug-cc-pVnZ basis sets, it is presumably the diffuse functions (which are optimised on the correlation energy) that increase the BSSE. However, within each basis set series the BSSE decreases from DZ to 6Z. Note that in the limit of using a hypothetical complete basis set on each fragment, the BSSE would be zero, as the fragments would not have a reason anymore to steal basis functions from the other fragment.

The failure of functionals like B3LYP to account for dispersion is well documented [25-40] and much effort has been devoted to develop functionals that describe dispersion effects. One strategy is to augment the density functional with an empirical dispersion term. This usually takes the form of a damped  $C_6/R^6$  term. Among the most popular of these techniques are the DFT-D methods of Grimme and co-workers [7,8,41]. Their latest dispersion add-on, D3, includes geometry-dependent dispersion coefficients and three-body contributions [41]. Recently, Grimme *et al.* developed a new composite density functional scheme, dubbed PBEh-3c, by coupling

an existing functional to atom-pairwise corrections for dispersion and BSSE [42].

Another strategy is to express the correlation energy as a fully nonlocal functional of the density or the orbitals. These include the Van der Waals density functionals originally introduced by Langreth *et al.* [43]. A different strategy is employed by the group of Truhlar. The Minnesota functionals are meta-GGA (generalised gradient approximation) functionals. They depend on many parameters, which are parameterised on high-level benchmark databases that include dispersion-bound systems. As a result, these functionals describe the short-to-medium range part of dispersion well, though the long-range dispersion is missed [44]. The M05 series was the first family of Minnesota functionals [3,45], which was followed by the M06 [4,5,46], M08 [47], M11 [48,49], M12 [50-52] and M15 [53,54] families. Parameter-free models have been proposed as well. For example, Becke and Johnson proposed a method that exploits the dipole moment of the exchange hole to generate dispersion interactions [55,56]. Dispersion correction is also important in solids; besides Grimme's D2 [8] and D3 [41] corrections, the empirical pairwise correction scheme of Tkatchenko and Scheffler [57] is also popular in calculations on crystals and surfaces and has been shown to commonly provide a useful prediction of structures and properties [58]. Reproducibility of results is extremely important for scientific credibility. Lejaeghere *et al.* demonstrated that most commonly-used codes and methods for DFT calculations on solids essentially give identical results [59]. The development and assessment of density functional theory methods for dispersion is a very active field, and an exhaustive review is beyond the scope of the current study. For reviews on dispersion-corrected DFT methods, we refer to Refs. [60] and [61].

In the current paper we study the Tyr-Gly book1 conformer at different levels of theory. Valdés *et al.* pointed out that a practical and cheap solution to the accurate study

of molecules with large anticipated intramolecular BSSE values is to apply DFT-D methodologies, as BSSE tends to be small in DFT calculations [62]. In the current study we employ B3LYP-D3. We also use the M06-2X functional, which was identified as being suitable for studying the folding of peptides for which dispersion is important [63]. In addition, we employ larger basis sets to reduce intramolecular BSSE.

## Methods

The B3LYP/6-31+G(d) and MP2/6-31+G(d) optimised structures of book1 were taken from Ref. [1]. The two structures mainly differ in the value of the  $\psi_{\text{tyr}}$  Ramachandran angle (the  $C_{\beta}(\text{Tyr})-C_{\alpha}(\text{Tyr})-C_{\text{carb}}(\text{Tyr})-N(\text{Gly})$  dihedral angle – see Figure 2 for atom labelling), which has a value of  $138^{\circ}$  in the B3LYP structure and  $120^{\circ}$  in the MP2 structure. Relaxed potential energy profiles were generated for rotation around the tyrosine  $C_{\alpha}-C_{\text{carb}}$  bond at various levels of theory, including B3LYP, B3LYP-D3, M06-2X and MP2, all with the 6-31+G(d) basis set. Step sizes between  $2^{\circ}$  and  $10^{\circ}$  were employed. We focused on the regions around  $138^{\circ}$  (where B3LYP has its minimum) and  $120^{\circ}$  (where MP2 has its minimum). In addition, a complete rotational scan with MP2 revealed another minimum at  $\sim 95^{\circ}$ ; we therefore also included the region around  $95^{\circ}$  in our investigation. Single-point calculations were performed for the M06-2X/6-31+G(d) optimised structures between  $90-110^{\circ}$  and  $110-180^{\circ}$  with MP2 and the cc-pVDZ and cc-pVTZ basis sets [64]. Single-point calculations were also performed for the B3LYP-D3/6-31+G(d) optimised structures between  $90-115^{\circ}$  with B3LYP-D3 and the cc-pVDZ, cc-pVTZ and cc-pVQZ [64] basis sets.

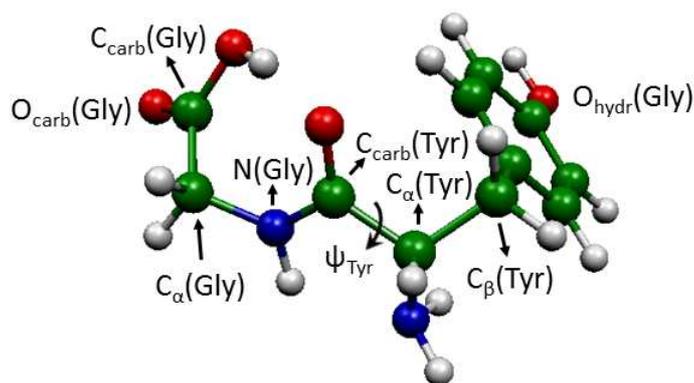


Figure 2. Atom labelling and definition of the  $\psi_{\text{Tyr}}$  torsion angle.

The minima at  $\sim 130^\circ$  and  $\sim 100^\circ$  were subsequently fully optimised with B3LYP, B3LYP-D3 and M06-2X using the 6-31+G(d) and cc-pVnZ ( $n = \text{D, T, Q}$ ) basis sets and with MP2 and the 6-31+G(d), cc-pVDZ and cc-pVTZ basis sets.

The intramolecular BSSE in selected MP2/6-31+G(d)-optimised conformers in the  $\psi_{\text{Tyr}}$  range of  $90\text{--}150^\circ$  was estimated by calculating the intermolecular BSSE in complexes consisting of phenol and *N*-formylglycine with the same conformations and spatial arrangements as in the partially optimised MP2 structures. This was done as follows [17]: The C(Tyr)H<sub>2</sub> and CH(Tyr)NH<sub>2</sub> groups were replaced by hydrogen atoms and the positions of the hydrogens were optimised in the resulting individual fragments (phenol and *N*-formylglycine) using MP2/6-31+G(d), keeping all other atoms fixed. The BSSE was then determined using Gaussian's Counterpoise keyword, employing the 6-31+G(d), cc-pVDZ, cc-pVTZ and aug-cc-pVTZ [64,65] basis sets. The same methodology was employed to estimate the BSSE in the fully-optimised MP2, B3LYP-D3 and M06-2X structures at  $\sim 100^\circ$ . In the B3LYP-D3 and M06-2X calculations, the positions of the replacement hydrogens were optimised with B3LYP/6-31+G(d) and M06-2X/6-31+G(d), respectively.

All calculations were performed using the Gaussian 09 suite of programs [66]. In all DFT calculations Gaussian's 'ultrafine' integration grid was employed. Note that

meta-functionals may yield spurious results when course or default integration grids are employed [67,68].

## Results

The relaxed energy scans for rotation around the tyrosine  $C_{\alpha}$ - $C_{carb}$  bond are displayed in Figure 3. The B3LYP profile exhibits only one minimum, at around  $140^{\circ}$ , whereas the MP2/6-31+G(d) profile shows two minima, at around  $120^{\circ}$  and  $95^{\circ}$ . The  $95^{\circ}$ -minimum is more stable than the  $120^{\circ}$ -minimum. The two MP2 minima differ in the orientation of the  $NH_2$  group, see Scheme 1. Type1 occurs in structures with the smaller  $\psi_{tyr}$  values (up to roughly  $110^{\circ}$ , depending on the method), whereas type2 occurs in the structures with  $\psi_{tyr}$  values above  $\sim 110^{\circ}$ . The MP2/6-31+G(d) profile shows that there are some angles around  $110^{\circ}$  where both types co-exist. In the B3LYP profile, on the other hand, there is a smooth transition from type1 to type2 from  $105^{\circ}$  to  $115^{\circ}$  (with eclipsed  $C_{\alpha}$  and amino hydrogens at  $110^{\circ}$ ). Type1 structures are more compact; the two amino acid residues are closer to each other. We note the close resemblance between the minimum at  $95^{\circ}$  and the previously studied book6 conformer optimised with MP2/6-31+G(d). The structures mainly differ in the orientation of the C-terminus, which in the book1 minimum at  $95^{\circ}$  is rotated such that the carboxylic OH-group can hydrogen bond with the carbonyl oxygen (Figure 4).

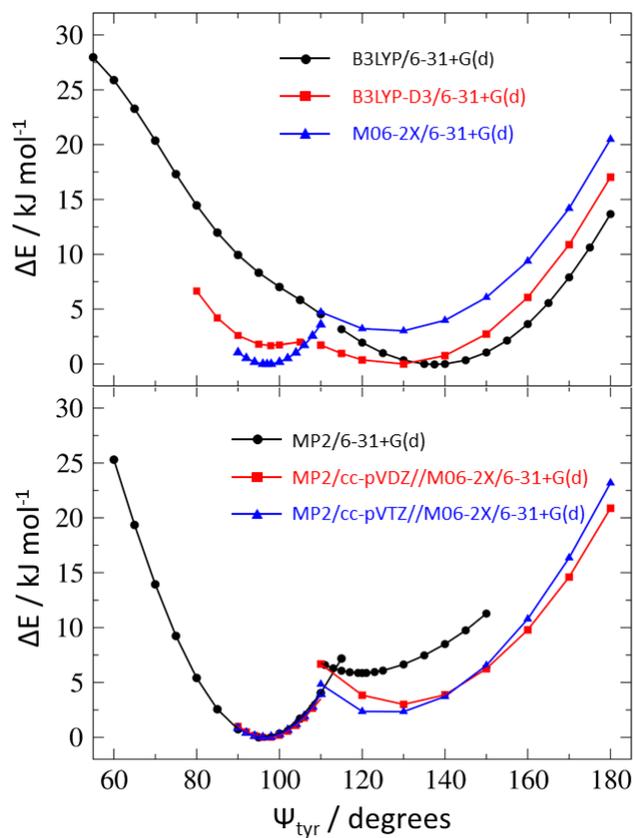
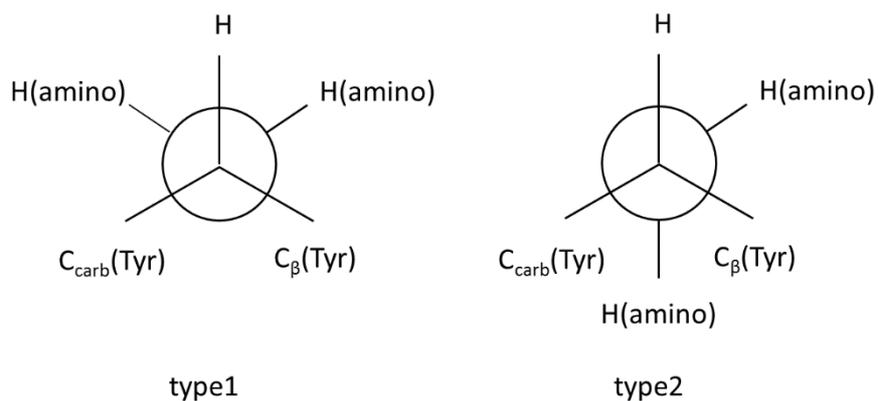


Figure 3. Potential energy profiles for rotation around the tyrosine  $C_{\alpha}$ - $C_{carb}$  bond, calculated at different levels of theory. The discontinuity in the curves indicate where type1 conformers convert into type2 conformers (see text).



Scheme 1. Newman projections showing the different  $NH_2$  orientation in type1 and type2 conformers.

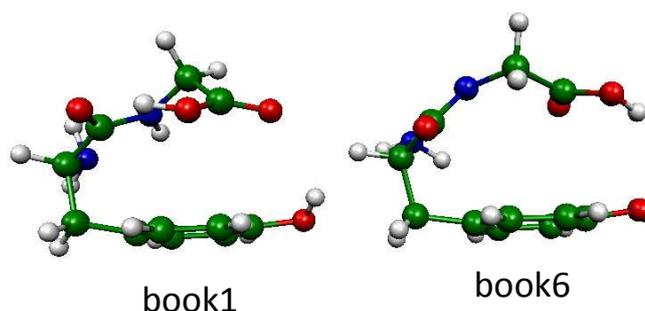


Figure 4. The MP2/6-31+G(d) optimised structures of the 95° book1 and the book6 conformers.

Adding a D3 term to the B3LYP calculations (see B3LYP-D3 profile) shifts the minimum at 130° to smaller torsion angle values. This confirms that missing dispersion in the B3LYP calculations is at least partially is to blame for the differing MP2 and B3LYP structures of this minimum. In addition, a minimum appears around 95°. This minimum is less stable than the minimum at larger  $\psi_{\text{tyr}}$  value. The M06-2X/6-31+G(d) profile also shows two minima, around 95° and around 130°, with the 95°-minimum lower in energy than the 130°-minimum. The absence of a B3LYP minimum around 95° could be due to missing dispersion in the B3LYP calculations. However, it may also be an artefact caused by intramolecular BSSE. To provide an estimate of the magnitude of the intramolecular BSSE in the MP2 calculations, we calculated the intermolecular BSSE for complexes consisting of phenol and *N*-formylglycine (see Methodology) at the MP2 level with different basis sets. Figure 5A shows the variation of the BSSE as a function of the  $\psi_{\text{tyr}}$  torsion angle. Type1 structures have clearly larger BSSE values than type2 structures. In both types, the BSSE reduces with increasing  $\psi_{\text{tyr}}$  values. Thus, the effect of BSSE is to artificially shift the minima towards lower  $\psi_{\text{tyr}}$  values.

The different NH<sub>2</sub> orientations in type1 and type2 conformers (see Scheme 1) lead to a slightly different conformation of the glycine residue, with the effect that the

two ends of the dipeptide are closer to each other in type1 structures. This can be quantified by the distance ( $R_{OO}$ ) between the glycine carboxylic oxygen,  $O_{carb}(Gly)$ , and the tyrosine hydroxyl oxygen,  $O_{hydr}(Tyr)$ . The inset in Figure 5A shows the variation of  $R_{OO}$  in the MP2/6-31+G(d) optimised structures as a function of  $\psi_{tyr}$ . The  $R_{OO}$  profiles remarkably resemble the BSSE curves. Clearly, shorter  $R_{OO}$  distances are correlated to increased BSSE.

We have estimated CP-corrected MP2/6-31+G(d) profiles by subtracting the intermolecular BSSE (calculated with MP2/6-31+G(d)) from the corresponding total energies. Figure 5B shows the resulting relative energies. The uncorrected curves (circles) correspond to those shown in Figure 3. The CP-corrected profiles clearly display their minimum at larger  $\psi_{tyr}$  values; this is particularly obvious for the type2 structures.

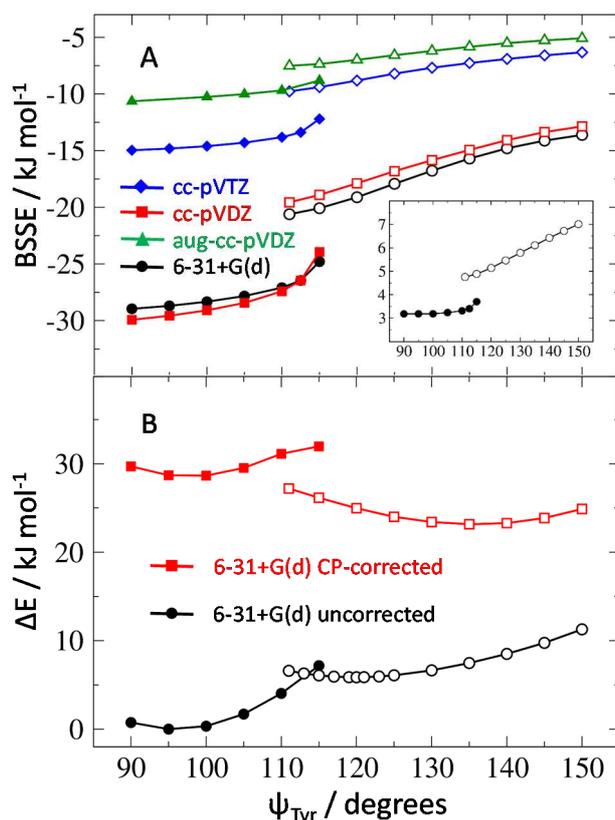


Figure 5. A. Variation of BSSE in the MP2 calculation as a function of  $\psi_{\text{Tyr}}$ . Closed symbols denote type1 structures; open symbols denote type2 structures. The inset shows the variation of the distance between O<sub>carb</sub>(Gly) and O<sub>hydr</sub>(Tyr) in the MP2/6-31+G(d) optimised structures as a function of  $\psi_{\text{Tyr}}$ . B. Uncorrected and CP-corrected MP2/6-31+G(d) potential energy profiles.

Table 1 lists the  $\psi_{\text{Tyr}}$  and R<sub>OO</sub> values of the two different book1 structures optimised at different levels of theory. Adding a D3 dispersion term to the B3LYP calculations decreases the  $\psi_{\text{Tyr}}$  torsion angle of the minimum at 130° from around 137 to around 128° (with cc-pVQZ), in close agreement with the M06-2X value calculated with the same basis set. Also listed in Table 1 is the BSSE for type1 structures calculated at different levels of theory. The BSSE is large even in the DFT calculations. This is in contrast to the results by Valdés *et al.*, who computed BSSE values of less than 2 kJ mol<sup>-1</sup> with B3LYP-D/TZVP for a benzene dimer complex modelling the interaction between the two aromatic rings in a folded conformer of the Phe-Tyr-Phe tripeptide (Phe = phenylalanine) [62]. Our value of 4.93 kJ mol<sup>-1</sup>, calculated with B3LYP-D3 and the slightly larger cc-pVTZ basis set, is considerably larger.

Interestingly, the 95°-minimum disappears at the B3LYP-D3 level when using the cc-pVQZ basis set. To investigate this further, we calculated single-point B3LYP-D3 energies with the cc-pVnZ ( $n = D, T, Q$ ) basis sets at the B3LYP-D3/6-31+G(d) structures optimised at fixed  $\psi_{\text{tyr}}$  torsion angles (Figure 6). While there is a clear barrier between the 100°- and 130° minima with the smaller basis sets, this vanishes at the cc-pVQZ level. This is presumably due to decreasing BSSE when increasing the basis set. As pointed out by Jensen, changes that are considered a ‘basis set effect’ are often at least partially attributable to intramolecular BSSE [69]. It is conceivable that further reduction of BSSE in the M06-2X and MP2 calculations (by further increasing the basis set size) would cause this minimum to vanish from the M06-2X and MP2 potential energy surfaces.

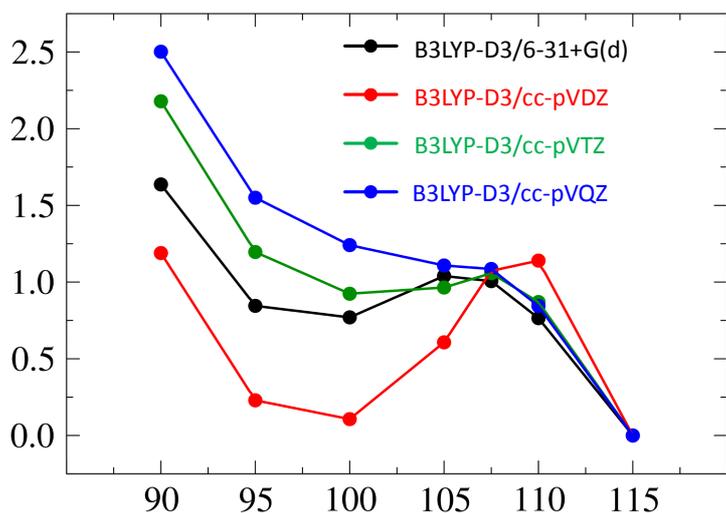


Figure 6. Single-point B3LYP-D3 calculations at the B3LYP-D3/6-31+G(d) optimised structures at fixed  $\psi_{\text{tyr}}$  torsion angles.

Table 1. Values of  $\psi_{\text{tyr}}$  (in degrees), BSSE (in  $\text{kJ mol}^{-1}$ ) and  $R_{\text{OO}}$  (in  $\text{\AA}$ ) for structures optimised at different levels of theory.

Method	minimum at 100°			minimum at 130°	
	$\psi_{\text{Tyr}}$	$R_{\text{OO}}$	BSSE <sup>a</sup>	$\psi_{\text{Tyr}}$	$R_{\text{OO}}$
B3LYP/6-31G(d,p)	n/a	n/a	n/a	137.4	6.08
B3LYP/cc-pVDZ	n/a	n/a	n/a	138.0	7.24
B3LYP/cc-pVTZ	n/a	n/a	n/a	136.8	7.33
B3LYP/cc-pVQZ	n/a	n/a	n/a	137.0	7.37
B3LYP-D3/6-31G(d,p)	97.9	4.27	4.28	128.4	6.25
B3LYP-D3/cc-pVDZ	97.7	3.06	19.27	131.1	6.33
B3LYP-D3/cc-pVTZ	100.6	4.66	4.93	128.8	6.32
B3LYP-D3/cc-pVQZ	n/a <sup>b</sup>	n/a <sup>b</sup>	n/a <sup>b</sup>	128.3	6.31
M06-2X/6-31G(d,p)	97.0	3.04	6.27	126.3	5.88
M06-2X/cc-pVDZ	97.5	2.95	18.08	130.6	6.24
M06-2X/cc-pVTZ	97.6	3.15	6.21	127.5	6.08
M06-2X/cc-pVQZ	98.3	3.62	2.36	127.6	6.10
MP2/6-31G(d,p)	95.9	3.19	28.63	119.8	5.12
MP2/cc-pVDZ	97.0	3.05	31.51	128.8	6.20
MP2/cc-pVTZ	96.6	3.28	14.50	124.1	5.71

<sup>a</sup> Converged to the minimum at 128°. <sup>b</sup> Estimated using BSSE values from complexes of *N*-formylglycine and phenol.

## Discussion and conclusions

The MP2/6-31+G(d) and B3LYP/6-31+G(d) levels of theory predict different structures for the Tyr-Gly conformer book1, with B3LYP yielding less folded structures as measured by the  $R_{\text{OO}}$  distance (B3LYP: 6.1  $\text{\AA}$ ; MP2: 5.1  $\text{\AA}$ ) and a larger  $\psi_{\text{tyr}}$  torsion angle (B3LYP: 137°; MP2: 120°). In the current work it is shown that missing dispersion in the B3LYP calculations as well as BSSE effects in the MP2 calculations are responsible for the different structures obtained by the MP2 and B3LYP calculations. This evidences the sensitivity to method and basis set size of geometries of molecules containing aromatic rings. To estimate the “correct” structure, we used DFT methods with inclusion of dispersion (M06-2X, B3LYP-D3) and large basis sets (cc-pVTZ and cc-pVQZ) in both DFT and MP2 calculations to minimise the intramolecular BSSE. The DFT methods yield  $\psi_{\text{tyr}}$  values of  $\sim 128^\circ$ . The BSSE is still sizable in the

MP2/cc-pVTZ calculations ( $\sim 8 \text{ kJ mol}^{-1}$ ), so the predicted  $\psi_{\text{tyr}}$  value of  $124^\circ$  is likely too small. Indeed, correcting the MP2/cc-pVTZ//MP2/6-31+G(d) profiles using MP2/cc-pVTZ BSSE values from phenol-*N*-formylglycine complexes yields a minimum around  $130^\circ$  (see Supporting Information), in excellent agreement with the B3LYP-D3/cc-pVQZ and M06-2X/cc-pVQZ results. We therefore estimate the ‘correct’  $\psi_{\text{tyr}}$  value to be  $\sim 130^\circ$ .

The MP2/6-31+G(d), M06-2X and B3LYP-D3 profiles show a separate minimum with  $\psi_{\text{tyr}}$  torsion angle around  $95^\circ$ . This second minimum has a different  $\text{NH}_2$  conformation and a more folded structure ( $R_{\text{OO}}$  around 3-4 Å, depending on the level of theory). However, when increasing the basis set in the B3LYP-D3 calculations from 6-31+G(d) to cc-pVQZ (thereby decreasing intramolecular BSSE) this minimum vanishes again. It therefore appears to be an artefact of large intramolecular BSSE effects. The minimum remains with M06-2X (up to cc-pVQZ basis set level) and with MP2 (with 6-31+G(d), cc-pVDZ and cc-pVTZ), but could presumably disappear when using larger basis sets. Although it is generally observed that BSSE is smaller in DFT than in MP2 calculations [70-72], there is recognition that BSSE may be sufficiently large in DFT calculations to warrant correction [16,73,74]. Our results support this. In addition, the results show that BSSE can generate artificial minima on the potential energy surface. We therefore recommend the use of large basis sets of preferably quadruple-zeta quality in MP2, DFT or DFT-D calculations in cases where intramolecular BSSE is expected to be large (*e.g.* molecules containing aromatic rings).

It would be good if we could verify our proposed book1 structure with experiment. However, a resonant two-photon ionisation (R2PI) study on gas-phase Tyr-Gly by de Vries *et al.*, which detected four distinct conformers, suggests that neither the hydroxyl nor the carboxylic OH groups are involved in hydrogen bonding, thereby

ruling out book1 [75]. The four observed conformers were assigned as those that were numbered 7, 10, 11 and 12 in our original study on Tyr-Gly [1]. The absence in the experiments of the most stable conformer predicted by computation therefore constitutes a discrepancy between theory and experiment. However, a subsequent computational study conducted in our group at the DSD-PBEP86-D3BJ/aug-cc-pVTZ level shows the importance of including thermal effects [76]. It is worth pointing out that, although the R2PI spectra are recorded at supercool temperatures, the sample is formed at high temperatures and may approximately preserve the Boltzmann distribution of its formation temperature. When free energy corrections at 400 K are added to the calculations, conformers with no hydrogen bonds become prevalent. At this temperature, a set of non-hydrogen-bonded conformers are predicted, including the folded book4 and book6 and two extended ones. Book4 and book6 were not originally assigned by de Vries *et al.* However, their presence is perfectly consistent with the experimental spectra, and a reassignment including book4 and book6 reconciles theory and experiment [76]. Unfortunately, book1 is not observed, probably because of the entropic bias against hydrogen bonding at high temperatures.

## Acknowledgements

We are grateful to EaStCHEM for computational support via the EaStCHEM Research Computing Facility. RH acknowledges funding from Higher Education Commission of Pakistan for her visit to the University of St Andrews as a visiting scholar.

## Disclosure statement

The authors declare no conflict of interest.

## References

- [1] D. Toroz, T. van Mourik, *Mol. Phys.* **104**, 559-570 (2006).
- [2] L.F. Holroyd, T. van Mourik, *Chem. Phys. Lett.* **442**, 42-46 (2007).

- [3] Y. Zhao, N.E. Schultz, D.G. Truhlar, *J. Chem. Theor. Comput.* **2**, 364-382 (2006).
- [4] Y. Zhao, D.G. Truhlar, *Theor. Chem. Acc.* **120**, 215-241 (2008).
- [5] Y. Zhao, D.G. Truhlar, *J. Chem. Phys.* **125**, 194101 (2006).
- [6] S. Grimme, T. Schwabe, *Phys. Chem. Chem. Phys.* **8**, 4398-4401 (2006).
- [7] S. Grimme, *J. Comput. Chem.* **25**, 1463-1473 (2004).
- [8] S. Grimme, *J. Comput. Chem.* **27**, 1787-1799 (2006).
- [9] T. van Mourik, *J. Chem. Theor. Comput.* **4**, 1610-1619 (2008).
- [10] A.E. Shields, T. van Mourik, *J. Phys. Chem. A* **111**, 13272 – 13277 (2007).
- [11] S.F. Boys, F. Bernardi, *Mol. Phys.* **19**, 553-566 (1970).
- [12] F. Jensen, *J. Chem. Phys.* **146**, 184109 (2017).
- [13] N.Y. Palermo, J. Csontos, M.C. Owen, R.F. Murphy, S. Lovas, *J. Comput. Chem.* **28**, 1208-1214 (2007).
- [14] T. van Mourik, *J. Comput. Chem.* **29**, 1-3 (2008).
- [15] J. Csontos, N.Y. Palermo, R.F. Murphy, S. Lovas, *J. Comput. Chem.* **29**, 4-7 (2008).
- [16] H. Kruse, S. Grimme, *J. Chem. Phys.* **136**, 154101 (2012).
- [17] T. van Mourik, P.G. Karamertzanis, S.L. Price, *J. Phys. Chem. A* **110**, 8-12 (2006).
- [18] D. Asturiol, M. Duran, P. Salvador, *J. Chem. Phys.* **128**, 144108 (2008).
- [19] D. Asturiol, M. Duran, P. Salvador, *J. Chem. Theor. Comput.* **5**, 2574-2581 (2009).
- [20] R.M. Balabin, *J. Chem. Phys.* **129**, 164202 (2008).
- [21] R.M. Balabin, *Mol. Phys.* **109**, 943-953 (2011).
- [22] F. Jensen, *J. Chem. Theor. Comput.* **6**, 100-106 (2010).
- [23] T. van Mourik, A.K. Wilson, K.A. Peterson, D.E. Woon, T.H. Dunning Jr., *Adv. Quantum Chem.* **31**, 105-135 (1999).
- [24] M. Gutowski, J.G.C.M.v.D.v.d. Rijdt, J.H.v. Lenthe, F.B.v. Duijneveldt, *J. Chem. Phys.* **98**, 4728-4737 (1993).
- [25] S. Kristyán, P. Pulay, *Chem. Phys. Lett.* **229**, 175-180 (1994).
- [26] P. Hobza, J. Šponer, T. Reschel, *J. Comput. Chem.* **16**, 1315-1325 (1995).
- [27] J.M. Pérez-Jordá, A.D. Becke, *Chem. Phys. Lett.* **233**, 134-137 (1995).
- [28] W. Kohn, Y. Meir, D.E. Makarov, *Phys. Rev. Lett.* **80**, 4153-4156 (1998).
- [29] S. Tsuzuki, T. Uchimaru, K. Tanabe, *Chem. Phys. Lett.* **287**, 202-208 (1998).
- [30] A. Millet, T. Korona, R. Moszynski, E. Kochanski, *J. Chem. Phys.* **111**, 7727-7735 (1999).
- [31] C. Tuma, A.D. Boese, N.C. Handy, *Phys. Chem. Chem. Phys.* **1**, 3939-3947 (1999).
- [32] A.K. Rappé, E.R. Bernstein, *J. Phys. Chem. A* **104**, 6117-6128 (2000).
- [33] N. Kurita, H. Sekino, *Chem. Phys. Lett.* **348**, 139-146 (2001).
- [34] S. Tsuzuki, H.P. Lüthi, *J. Chem. Phys.* **114**, 3949-3957 (2001).
- [35] T. van Mourik, R.J. Gdanitz, *J. Chem. Phys.* **116**, 9620-9623 (2002).
- [36] E.R. Johnson, R.A. Wolkow, G.A. DiLabio, *Chem. Phys. Lett.* **394**, 334-338 (2004).
- [37] T. van Mourik, *Chem. Phys.* **304**, 317-319 (2004).
- [38] R. Crespo-Otero, L.A. Montero, W.-D. Stohrer, J.M. García de la Vega, *J. Chem. Phys.* **123**, 134107 (2005).
- [39] S.M. Cybulski, C.E. Severson, *J. Chem. Phys.* **122**, 014117 (2005).
- [40] I. Dąbkowska, P. Jurecka, P. Hobza, *J. Chem. Phys.* **122**, 204322 (2005).

- [41] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J. Chem. Phys.* **132**, 154104 (154119 pages) (2010).
- [42] S. Grimme, J.G. Brandenburg, C. Bannwarth, A. Hansen, *J. Chem. Phys.* **143**, 054107 (2015).
- [43] M. Dion, H. Rydberg, E. Schröder, D.C. Langreth, B.I. Lundqvist, *Phys. Rev. Lett.* **92**, 246401 (2004).
- [44] L. Goerigk, H. Kruse, S. Grimme, *Chem. Phys. Chem.* **12**, 3421-3433 (2011).
- [45] Y. Zhao, N.E. Schultz, D.G. Truhlar, *J. Chem. Phys.* **123**, 161103 (2005).
- [46] Y. Zhao, D.G. Truhlar, *J. Phys. Chem. A* **110**, 13126 - 13130 (2006).
- [47] Y. Zhao, D.G. Truhlar, *J. Chem. Theor. Comput.* **4**, 1849-1868 (2008).
- [48] R. Peverati, D.G. Truhlar, *J. Phys. Chem. Lett.* **2**, 2810-2817 (2011).
- [49] R. Peverati, D.G. Truhlar, *J. Phys. Chem. Lett.* **3**, 117-124 (2012).
- [50] R. Peverati, D.G. Truhlar, *J. Chem. Theor. Comput.* **8**, 2310-2319 (2012).
- [51] R. Peverati, D.G. Truhlar, *Phys. Chem. Chem. Phys.* **14**, 13171-13174 (2012).
- [52] R. Peverati, D.G. Truhlar, *Phys. Chem. Chem. Phys.* **14**, 16187-16191 (2012).
- [53] H.S. Yu, X. He, D.G. Truhlar, *J. Chem. Theor. Comput.* **12**, 1280-1293 (2016).
- [54] H.S. Yu, X. He, S.L. Li, D.G. Truhlar, *Chem. Sci.* **7**, 5032-5051 (2016).
- [55] A.D. Becke, E.R. Johnson, *J. Chem. Phys.* **122**, 154104 (2005).
- [56] E.R. Johnson, A.D. Becke, *J. Chem. Phys.* **123**, 024101 (2005).
- [57] A. Tkatchenko, M. Scheffler, *phys. Rev. Lett.* **102**, 073005 (2009).
- [58] L. Kronik, A. Tkatchenko, *Acc. Chem. Res.* **47**, 3208-3216 (2014).
- [59] K. Lejaeghere, G. Bihlmayer, T. Björkman, P. Blaha, S. Blügel, V. Blum, D. Caliste, I.E. Castelli, S.J. Clark, A. Dal Corso, S. de Gironcoli, T. Deutsch, J.K. Dewhurst, I. Di Marco, C. Draxl, M. Dułak, O. Eriksson, J.A. Flores-Livas, K.F. Garrity, L. Genovese, P. Giannozzi, M. Giantomassi, S. Goedecker, X. Gonze, O. Grånäs, E.K.U. Gross, A. Gulans, F. Gygi, D.R. Hamann, P.J. Hasnip, N.A.W. Holzwarth, D. Iușan, D.B. Jochym, F. Jollet, D. Jones, G. Kresse, K. Koepnik, E. Küçükbenli, Y.O. Kvashnin, I.L.M. Locht, S. Lubeck, M. Marsman, N. Marzari, U. Nitzsche, L. Nordström, T. Ozaki, L. Paulatto, C.J. Pickard, W. Poelmans, M.I.J. Probert, K. Refson, M. Richter, G.-M. Rignanese, S. Saha, M. Scheffler, M. Schlipf, K. Schwarz, S. Sharma, F. Tavazza, P. Thunström, A. Tkatchenko, M. Torrent, D. Vanderbilt, M.J. van Setten, V. Van Speybroeck, J.M. Wills, J.R. Yates, G.-X. Zhang, S. Cottenier, *Science* **351** (2016).
- [60] S. Grimme, *Wiley Interdisciplinary Reviews: Computational Molecular Science* **1**, 211-228 (2011).
- [61] J. Klimeš, A. Michaelides, *J. Chem. Phys.* **137**, 120901 (2012).
- [62] H. Valdés, V. Klusák, M. Pitoňák, O. Exner, I. Starý, P. Hobza, L. Rulíšek, *J. Comput. Chem.* **29**, 861-870 (2008).
- [63] E. Gloaguen, B. de Courcy, J.P. Piquemal, J. Pilmé, O. Parisel, R. Pollet, H.S. Biswal, F. Piuzzi, B. Tardivel, M. Broquier, M. Mons, *J. Am. Chem. Soc.* **132**, 11860-11863 (2010).
- [64] T.H. Dunning Jr., *J. Chem. Phys.* **90**, 1007 (1989).
- [65] R.A. Kendall, T.H. Dunning Jr., R.J. Harrison, *J. Chem. Phys.* **96**, 6796 (1992).
- [66] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G.A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H.P. Hratchian, A.F. Izmaylov, J. Bloino, G. Zheng, J.L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. Montgomery, J. A., J.E. Peralta, F. Ogliaro, M. Bearpark, J.J. Heyd, E. Brothers,

- K.N. Kudin, V.N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J.M. Millam, M. Klene, J.E. Knox, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, R.L. Martin, K. Morokuma, V.G. Zakrzewski, G.A. Voth, P. Salvador, J.J. Dannenberg, S. Dapprich, A.D. Daniels, Ö. Farkas, J.B. Foresman, J.V. Ortiz, J. Cioslowski, D.J. Fox, Gaussian 09. Gaussian, Inc., Wallingford, CT, 2009.
- [67] S.E. Wheeler, K.N. Houk, *Journal of Chemical Theory and Computation* **6**, 395-404 (2010).
- [68] I.D. Mackie, G.A. DiLabio, *Phys. Chem. Chem. Phys.* **12**, 6092-6098 (2010).
- [69] F. Jensen, *Chem. Phys. Lett.* **261**, 633-636 (1996).
- [70] M.C. Daza, J.A. Dobado, J.M. Molina, P. Salvador, M. Duran, J.L. Villaveces, *J. Chem. Phys.* **110**, 11806-11813 (1999).
- [71] Y.-K. Han, K.H. Kim, S.-K. Son, Y.S. Lee, *Bull. Korean Chem. Soc.* **23**, 1267-1271 (2002).
- [72] S. Simon, M. Duran, J.J. Dannenberg, *J. Phys. Chem. A* **103**, 1640-1643 (1999).
- [73] J.G. Brandenburg, M. Alessio, B. Civalleri, M.F. Peintinger, T. Bredow, S. Grimme, *J. Phys. Chem. A* **117**, 9282-9292 (2013).
- [74] D. Escudero, A. Frontera, D. Quiñonero, P.M. Deyà, *Chem. Phys. Lett.* **455**, 325-330 (2008).
- [75] A. Abo-Riziq, L. Grace, B. Crews, M.P. Callahan, T. van Mourik, M.S. de Vries, *J. Phys. Chem. A* **115**, 6077-6087 (2011).
- [76] L.F. Holroyd, T. van Mourik, *Chem. Phys. Lett.* **621**, 124-128 (2015).